HAND-HELD OR HANDS-FREE? THE EFFECTS OF WIRELESS PHONE INTERFACE TYPE ON PHONE TASK PERFORMANCE AND DRIVER PREFERENCE

Elizabeth N. Mazzae, MSE
National Highway Traffic Safety Administration
Vehicle Research and Test Center
East Liberty, OH

Thomas A. Ranney, Ph.D.
Transportation Research Center Inc.
East Liberty, OH

Ginger S. Watson, Ph.D. and
Judith A. Wightman, MA
University of Iowa, National Advanced Driving Simulator,
Iowa City, IA

The number of wireless phone subscribers in the U.S. is constantly growing. Studies have shown that use of wireless phones while driving contributes to crashes. Efforts to pass legislation allowing only hands-free wireless phone use while driving are widespread and based on the assumption that a hands-free interface is safer than a hand-held one. The National Highway Traffic Safety Administration conducted a driving simulator study to examine the effects of phone interface type on driving performance and drivers' ability to perform phone tasks. Participants' phone interface preferences were also recorded. Fifty-four participants drove a simulated freeway route with each of three phone interfaces: hand-held, headset hands-free (voice dialing and headset), and voice dialing hands-free (voice dialing with cradle-mounted speaker phone). Although post-drive questionnaire results show that participants rated the hand-held interface to be most difficult to use, this interface was associated with the fewest dialing errors and significantly faster dialing times than the two hands-free interfaces. Drivers answered the phone more quickly when using the voice dialing hands-free phone interface than when using the hand-held or headset hands-free interfaces. Younger drivers answered the phone significantly more quickly than the older drivers. Hang up times were significantly faster in the voice dialing hands-free condition and were slowest in the headset hands-free condition. Conversation task performance did not differ as a function of phone interface. Thus, although drivers considered them easier to use than hand-held phone interfaces, hands-free interfaces were more time-consuming to interact with while driving. Further analyses of these data will investigate how other aspects of phone task performance are related to eye glance behavior and driving performance.

INTRODUCTION

As of May, 2004 there were over 163 million wireless phone subscribers (CTIA, 2004) in the U.S. and the number is constantly growing. A substantial portion of this group uses their wireless phones while driving, at least occasionally. According to a survey by the National Highway Traffic Safety Administration (NHTSA), at any given time, an estimated 3 percent of those driving passenger vehicles on America's roadways are talking on hand-held wireless phones (NHTSA, 2001). The 2000 Motor Vehicle Occupant Safety Survey estimated that 73 percent of drivers who said they usually have a wireless phone in their vehicle with them use a hand-held phone, while 22 percent use hands-free equipment (NHTSA, 2001).

Studies have shown that use of wireless phones while driving contributes to crashes (NHTSA, 1997). The crash-related effects of wireless phone use while driving is a controversial issue, and has been under public scrutiny in recent years. Across the U.S. and in other countries, numerous efforts are underway to pass legislation that allows only hands-free wireless phone use while driving. This move is based on the assumption that any technology that reduces the visual-manual demands of wireless phone use must be safer, since the driver can keep both hands on the wheel and both eyes on the road when using a hands-free system. It is interesting to note that hands-free wireless phones most commonly allow only for hands-free conversation; accessing the phone, dialing, and hanging up still involve physical manipulation of the phone as well as directed glances toward it. The legislative initiatives reflect this level of technology that involves manual dialing. In theory, voice dialing should relieve the visual demand related to dialing a wireless phone while driving. However, if the voice recognition capability of the hands-free interface is poor, placing a call may actually take longer than it would with manual dialing, thus increasing the duration in which the driver interacts with the phone and allocates mental resources toward placing a call.

NHTSA conducted a study using the National Advanced Driving Simulator (NADS) to examine the effects of wireless phone interface type on driving performance and phone task performance. The study also investigated driver preferences regarding phone interfaces. This paper discusses drivers' preferences regarding phone interfaces and compares them to objective phone use data. A subsequent NHTSA report will further discuss these results along with related driving performance results.

METHOD

Approach

An experiment was conducted in which participants drove a 17-mile freeway scenario on the NADS while performing phone tasks. The scenario consisted of straight segments of divided highway with moderate traffic and interchanges that required exiting and merging behavior. This route was driven once for each phone interface. Simulated phone conversations were staged to coincide with selected driving situations to ensure that all participants used the phone under comparable driving conditions.

Participants

A total of 54 licensed drivers (27 males and 27 females) participated. The participants were equally divided among three age groups: Younger (18-25), Middle (30-45), and Older (50-60). Persons eligible to participate had owned a wireless phone for a minimum of six months and reported using a wireless phone while driving at least once per day.

Experimental Design

The experiment had three factors: wireless phone interface (within-subjects), phone call status (within-subjects), and driver age group. The three levels of wireless phone interface included hand-held (HH) with manual dialing, headset with voice digit dialing (headset hands-free, or HHF), and hands-free speaker kit with voice digit dialing and fixed phone location using a console-mounted cradle (voice dialing hands-free, or VDHF). Levels of phone call status included incoming, outgoing, and no call (baseline). Age groups included Younger, Middle, and Older, as indicated above.

Dependent measures of phone task performance included dialing time, answering time, conversation task performance, hang up time, and opinions about each wireless phone interface. The dialing time began when the call placement was requested and ended when the start of the conversation task was heard. Answering time was defined as the time from when the first ring was heard in the cab until the start of the conversation task was heard. Conversation task performance was measured in terms of the number of correct judgments and the number of correctly recalled words. Hang up time began when the word, "stop" was heard at the completion of the conversation period and ended when the person returned their hands to the steering wheel.

A 24-item post-drive questionnaire was administered to collect participant attitudes, preferences, and previous experience with each type of wireless phone interface. Specific questions focused on perceived ease or difficulty in dialing, conversing, and driving with each interface, preference for each interface method, and overall preference for a manual vs. voice dialing method.

Phone Conversation Task (Baddeley Task)

The conversation task was a modification of the Baddeley working memory span task (Baddeley, Logie, and Nimmo-Smith, 1985). Participants were required to listen to sentences and determine whether or not they made sense (decision-making/judgment component). All sentences were of the following construction: Subject - action verb – object, and used common English words. Sentences were constructed so that the judgment task (sensical/nonsensical) could not be made until the object was heard. This required the participant to pay attention to the entire sentence before answering and also forced a consistent start to the response period (i.e., the completion of the last word of a sentence was the beginning of the response period.). After each group of four sentences, participants were required to recall either the subject or object of each sentence in the group (memory recall component). The recall components (subject/object) were balanced across calls such that half requested recall of the sentence subjects and half requested recall of the sentence objects.

Before each group of sentences, participants were told to recall either the subjects or objects of the sentences. Each group of four sentences comprised one Baddeley task trial. To ensure that phone calls spanned the desired events, calls were constructed to last approximately 3.5 minutes. Thus, each call had six groups of four for a total of 24 sentences (12 meaningful + 12 nonsensical sentences).

Monetary Incentives

Participants were paid a base pay amount of $30, plus incentive pay. The monetary incentives were used to establish priorities for participants and to promote a balance between primary (driving) and secondary (phone communication) task performance. The incentive scheme was intended as a method of reliably rewarding participants for performance; it was not intended or designed for use as a dependent measure.

The monetary rewards and penalties were based on a total number of points allocated for each task during each drive. Money was earned for driving safely and attentively and for completing phone tasks accurately and quickly. Unsafe driving, including speeding, reckless driving, and collisions that could have been avoided, resulted in monetary penalties. For example, extreme steering responses or excessively hard braking were considered unsafe responses. Participants started with a specified number of points and then lost points for not performing well and/or gained points for performing above expectations. It was not possible for participants to "lose" money beyond what was allocated for incentives (i.e., no pay was taken away from the base pay). Incentive pay ranged between $0 and $8.00 per drive.

Apparatus

The NADS, which is owned by the NHTSA and operated by the University of Iowa, consists of a vehicle cab attached to a turntable inside a 24-foot projection dome. The dome is mounted on a six-legged motion platform that moves about a 64-foot (x) by 64-foot (y) bay, allowing motions such as lane changing, spinning, and skidding. Control feel systems provide a realistic feel of the road and vehicle response. The visual system provides a 360-degree field of view, and a surround-sound system provides sounds coordinated with the other sensory systems. Every dynamic aspect of the virtual environment can be specified, including weather conditions, vehicle interactions, and sudden mechanical failures.

A Samsung model A460 wireless flip phone was used in all three interface conditions. To promote realism, actual digital PCS wireless phone service was used, rather than simulating the phone calls. Using actual wireless phone service allowed the use of an unmodified, commercially available wireless phone and required only adding a digital signal repeater to transmit the signal into the NADS dome. Voice digit dialing was accomplished using the Sprint PCS Voice Command System.

Procedure

Prior to driving, each participant received instruction and training on the use of the three interface types, as well as the conversation task, and route features.

Each participant completed a single session lasting 3 hours. During this session, the participant drove the same scenario route three times, once for each phone interface. The order of presentation of phone interface conditions was randomized. Each traversal of the route involved one incoming and one outgoing call. The order of presentation of incoming and outgoing calls was balanced.

The route consisted of a four-lane divided freeway with a 65-mph speed limit with traffic present. The route generally consisted of four straight segments of nearly equal length joined by right-side interchanges requiring exiting and merging behavior. The treatment drives were approximately 15 minutes in length and required participants to drive three segments of the divided freeway route. The route segments corresponded, respectively, to the incoming phone call, outgoing phone call, and baseline (no call) periods. Each route segment involved a series of interactions between the driver and the scenario vehicles (i.e., events). Events included a sudden lead-vehicle cut-in (LV cut-in), sudden braking by the lead vehicle (LV brake), and a car following event. Each traversal of the route was associated with a different order of events. The intention of the scenario design was to overlap the events with the 3.5-minute conversation task periods. Although the conversation task component of each call was presented continuously, the conversation task period was separated into three consecutive intervals based on the associated driving tasks: a car-following segment of 60 seconds (during which measures of the participant's ability to accurately follow the speed changes of a lead vehicle were obtained), a 30-second segment during which a discrete event (LV cut-in or LV brake) occurred, and a 45-second merging segment. Overall, 40 percent of the scenario involved phone task performance coupled with scenario events while 18.3 percent of the scenario consisted of baseline driving in which participants experienced scenario events while they were not using the phone. The remaining 41.7 percent of the scenario involved uneventful driving.

Incoming calls, placed by a remote experimenter, were answered by the participant as he/she was trained to do. For the HH and HHF interfaces, four manual steps were involved in receiving a call (pick up the phone, extend the antenna, flip the phone open, press 'talk' button). For the VDHF interface, receiving a call involved pressing the 'talk' button on the cradle-mounted phone.

For outgoing calls, the in-vehicle experimenter stated, "Please call (Angie, Dan, Ellen…) now," when prompted by a remote experimenter upon reaching a specific geographical point in the scenario database. Upon hearing this request, the participant reached for the phone to dial the appropriate number from a list mounted on the dashboard. For the HH interface, five manual steps were involved in placing a call (same as for call receipt with the addition of a dialing step). For the HHF and VDHF interfaces, placing a call required the driver to press '*' then 'talk' and then say "Call 319-335-xxxx," followed by saying "yes" or "no" to verbally indicate whether the system had correctly understood the number.

After all driving was completed the post-drive questionnaire was administered.

RESULTS

Questionnaire results were analyzed to determine the participants' preferences among the three wireless phone interface conditions. The pattern of preferences was compared with performance differences for initiating and terminating phone calls, as well as for performance on the conversation task while driving. All inferential analyses were conducted using the SAS System for Windows.

Questionnaire Results

Participants rated various aspects of wireless phone use while driving including ease of dialing and conversing, as well as ease of maintaining lane position and speed. Table 1 summarizes these results.

Table 1. Selected questionnaire results by phone interface

QUESTION Scale: 1= Extremely Easy 2= Easy 3= Slightly Easy 4= Neutral 5= Slightly Difficult 6= Difficult 7= Extremely Hard	**PHONE INTERFACE**			
	HH	**HHF**	**VDHF**	**Personal Phone**
Dialing while driving	5.0	3.3	2.9	3.8
Staying in lane while dialing	5.1	3.4	2.9	3.3
Maintaining speed limit while dialing	4.9	3.7	3.3	3.0
Carrying on conversation while driving	3.5	2.7	2.6	2.2
Staying in lane while talking	3.4	2.8	2.6	2.4
Maintaining speed limit while talking	4.4	3.3	3.2	2.6
Hearing/Understanding conversation	3.3	3.0	2.7	2.8
Overall Average	4.2	3.2	2.9	2.9

The HH interface was rated to be most difficult to use while driving for all conditions probed, while the VDHF interface was considered the easiest. Participants generally rated their personal phones to be somewhat easier to use than the HH interface and somewhat more difficult to use than the VDHF interface. Taking an average of the 7 ratings reveals the VDHF interface as the one participants considered the easiest to use and comparable to participants' personal phones, followed by the HHF and HH interfaces, respectively.

Questionnaire results showed that participants' average length of use of a wireless phone was 6.01 years (MIN=0.5 years, MAX=30 years, SD=4.7 years). Nineteen participants reported owning hands-free kits. Eight participants did not know if they had a hands-free kit. Twelve participants did not know whether their personal wireless phone had a speakerphone function. Five of the 54 participants reported regularly using a headset with their wireless phone. Three participants reported regularly using voice tag dialing, while two participants reported regularly using voice digit dialing.

Phone Task Performance Results

Phone Dialing Time. Overall, approximately 30% of the outgoing phone calls required more than one attempt. Use of the hand-held interface was least likely to require additional attempts (18%), while the Headset interface was most likely to require additional attempts (40%). This difference was statistically significant ($z = 2.16$, p = .0158). The percentage of VDHF calls requiring more than one attempt was 31%, which was not statistically different from either of the other interface values. An ANOVA was computed on dialing times with interface and age group as the independent variables. This analysis found significant main effects of interface, $F(2,32) = 56.34$, $p < .0001$, and age group, $F(2,39) = 4.94$, $p = 0.0122$. The interaction of these two factors was also statistically significant, $F(4,32) = 4.23$, $p = 0.0073$. This effect is shown in Figure 1.

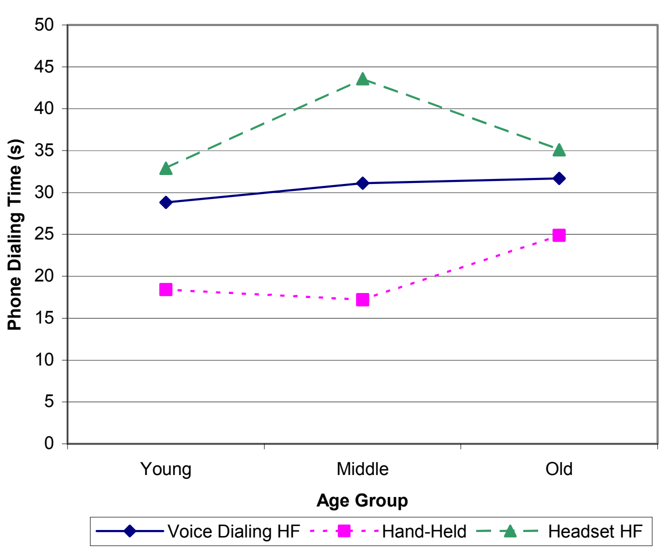

Figure 1. Dialing time by phone interface and age group

Examination of differences between pairs of means revealed that the hand-held interface was associated with significantly faster dialing times than the other two interfaces for all three age groups. However, dialing times for the hands-free and headset interfaces were significantly different only for the middle age group. The significant main effect of age group reflects the fact that when dialing times are collapsed across interface condition, the mean time for the older group is significantly slower than for the other two groups.

Phone Answering Time. To compare answering time for calls made with each interface, an ANOVA was computed using interface and age group as the independent variables. Significant main effects were found for interface condition, $F(2,101) = 128.16$, $p < .0001$, and age group, $F(2,51) = 8.32$, $p = 0.0007$. Examination of the means revealed that drivers answered much more quickly when using the VDHF interface ($M = 5.25$ s) than when using the HH ($M = 10.28$ s) or the HHF ($M = 10.25$ s) interfaces. The main effect of age group reflects the fact that the younger drivers ($M = 7.9$ s) answered significantly more quickly than the older drivers ($M = 9.7$ s). The middle age group had an intermediate value ($M = 8.2$ s) that was not statistically different from either group. The Interface x Age Group interaction was not significant, $F(4,101) = 2.04$, $p = 0.0947$.

Phone Hang Up Time. An ANOVA for phone hang up time was computed using the same model as described above. Interface condition was found to have a significant main effect, $F(2,101) = 96.1$, $p < .0001$. Specifically, hang up times were faster in the VDHF condition ($M = 4.5$ s), slowest in the HHF condition ($M = 11.0$ s), with an intermediate value in the HH condition ($M = 9.0$ s). Differences between all pairs of means were statistically significant.

Phone Conversation Task Performance. Analyses were conducted to examine the effects of phone interface, age, and gender on conversation (Baddeley) task scores for the six calls. Two phone task performance measures were considered: overall judgment (total number of sentences correctly identified as sensible or nonsensical), and overall recall (total number of key words correctly recalled). For each call, scores on each measure could range from 0 to 24. An alpha level of .05 was used for all statistical tests, and the Tukey HSD procedure was used for all follow-up tests.

For cases in which a restart of the simulator resulted in a particular call or portion of a call being heard more than once, phone task data from the final playing of the call were used in order to allow for comparison to the corresponding engineering data. For cases in which issues beyond the participants' control caused some sentences within a call to be missed, when four or more sentences (out of 24) were missed, these means scores were entered as missing. Thus, the general linear model (GLM) procedure was used instead of the ANOVA procedure because the former is less sensitive to the effects of unequal cell sizes.

First, a model using age and gender was examined. For judgment, significant differences were found for age, F

(2, 284) = 4.96, $p < .01$. Follow-up tests revealed that the older group correctly judged significantly fewer words (M = 23.06) than the middle (M = 23.49) and young (M = 23.41) groups. Effects of gender were not significant. For recall, significant differences were found for age, F (2, 302) = 53.38, $p < 0.01$. Follow-up tests revealed that the young group recalled significantly more words on average (M = 20.87) than the middle (M = 17.79) and older (M = 15.89) groups, and the middle group recalled more words on average than the older group. The effects of gender, and the interaction of age and gender were not significant.

The second model examined the effects of age, gender, and phone interface on conversation task performance. Age was again significant for both judgment and recall. Effects of gender, phone interface, and the interaction of age, gender, and phone interface were not significant.

DISCUSSION

Although participants rated the hand-held interface to be most difficult to use, this interface was associated with the fewest dialing errors (in terms of the number of attempts per dialing trial). Participants' feelings that the HH interface was the most difficult to use were also not supported by dialing time results, which showed that the HH interface was associated with significantly faster dialing times than the other two interfaces for all three age groups. Shorter dialing times for the HH interface may be attributable to participants' prior experience with hand-held wireless phones, which was approximately 6 years on average. Furthermore, participants were likely to have had experience using traditional push-button phones for home use, whereas it is likely that far fewer people had previous experience with voice-based dialing. However, it should be noted that the length of time required to perform voice digit dialing depends on the interface being used. This study used the Sprint PCS Voice Command system, since it was assumed that a system-based voice dialing interface would be more likely to have better voice recognition capability than phone-based voice dialing. Some newer model phones feature integrated voice digit dialing capability, which may allow shorter dialing times than the type used in this study. Use of voice "tags" for dialing may also afford shorter dialing times; however, voice digit dialing was chosen for implementation in this study since it provided the most direct comparison between manual and voice dialing.

The VDHF interface was associated with significantly faster answering times than those for the other two interfaces, which were approximately equal. Hang up times followed the same general trend with the VDHF condition (M = 4.5 s) being the fastest, and HH and HHF being slower at 9 s and 11 s, respectively. For both answering and hanging up the phone, one would expect similar task durations for the HH and HHF phone interfaces since the same basic steps were required to initiate or end the call (i.e., grab phone, flip phone open, extend antenna, press talk). Answering and hanging up using the VDHF interface each required a single button press.

Conversation task performance did not differ as a function of phone interface. Age was the only examined variable significantly related to phone task performance, with younger individuals performing better than older individuals. Gender was unrelated to phone task performance.

CONCLUSIONS

This study examined the phone interface related preferences of participants of three age groups and contrasted these with phone use performance as a function of phone interface. Results showed that in most cases participants overestimated the ease of use afforded by hands-free phone interfaces. In general, participants considered the hand-held interface to be most difficult to use, followed by the headset hands-free and voice dialing hands-free interfaces, respectively. However, significant differences among interfaces were evident for dialing and hanging up. The hand-held interface was associated with the fewest dialing errors and significantly faster dialing times than the two hands-free interfaces for all three age groups. Findings concerning the time taken to dial and answer are directly applicable to real world driving since a real phone connection was used in the study. No differences were found among interface conditions in phone conversation task performance, including judgments about the sentences and recall of sentence subjects or objects.

Further analyses of these data will investigate how duration and other aspects of phone task performance are related to eye glance behavior and driving performance in both uneventful and conflict driving situations. These and other additional details on this research will be available in upcoming NHTSA reports.

ACKNOWLEDGEMENTS

The authors acknowledge the technical support of Yiannis Papelis, Omar Ahmad, and Julie Qidwai of the University of Iowa.

REFERENCES

CTIA's World of Wireless Communications, Retrieved May 3, 2004 1:00 PM EDT. http://www.ctia.org. Cellular Telecommunications & Internet Association. "163,227,785 Current US Wireless Subscribers."

National Highway Traffic Safety Administration (2001). Passenger Vehicle Driver Cell Phone Use Results from the Call 2000 National Occupant Protection Use Survey. Research Note No. DOT HS 809 293. Washington, D.C: U.S. Department of Transportation.

National Highway Traffic Safety Administration (1997). An Investigation of the Safety Implications of Wireless Communications in Vehicles. Technical Report No. DOT HS 808 635. Washington, D.C: U.S. Department of Transportation.

Baddeley, A., Logie, R., and Nimmo-Smith, I. (1985). Components of fluent reading. Journal of Memory and Language, 24, 119-131.

Brookhuis, K., Waard, D. d., and Mulder, B. (1994). Measuring driving performance by car-following in traffic. Ergonomics, 37, 427-434.