ELSEVIER

Accident Analysis and Prevention 40 (2008) 675–682

ACCIDENT ANALYSIS & PREVENTION

www.elsevier.com/locate/aap

Assessing the awareness of performance decrements in distracted drivers

William J. Horrey *, Mary F. Lesch, Angela Garabet

Liberty Mutual Research Institute for Safety, 71 Frankland Road, Hopkinton, MA 01748, USA

Received 30 April 2007; accepted 7 September 2007

Abstract

Many studies have documented the performance decrements associated with driver distractions; however, few have examined drivers' awareness of these distraction effects. The current study measured how well-calibrated drivers are with respect to performance decrements from distracting tasks. In this test track study, 40 younger and older drivers completed a series of tasks on a hand-held or hands-free cell phone while driving around a course in an instrumented vehicle. Subjective estimates of performance decrements were compared to actual performance decrements. Although their driving performance suffered in dual-task conditions, drivers were generally not well-calibrated to the magnitude of the distraction effects ($r = -.38$ to .16). In some cases, estimates of distraction were opposite of the observed effects (i.e., smaller estimates of distraction corresponded to larger performance deficits). Errors in calibration were unassociated with several measures of overconfidence in safety and skill, among other variables. We discuss the implications of these findings for potential mitigation strategies for distracted driving.

Keywords: Driver distraction; Awareness of performance; Confidence; Younger and older drivers

1. Introduction

Over the past several years, there have been numerous studies documenting the effects of engaging in cell phone conversations or other distracting tasks while driving. In general, these studies have shown that distracted drivers have slowed responses to critical traffic events or to discrete stimuli and are more likely to miss external events such as a changing traffic light, among other effects (Alm and Nilsson, 1994; Hancock et al., 2003; McKnight and McKnight, 1993; Strayer and Johnston, 2001).

To date, much of the focus has been on the presence and magnitude of performance decrements for different types of in-vehicle tasks or device configurations. These efforts may help in legislative initiatives and design recommendations. For example, many states have enacted or proposed legislation to ban certain in-vehicle devices such as hand-held cellular phones (Sundeen, 2005). Most legislative bans target hand-held cell phones; however, many studies have shown that hands-free cell phone conversations are just as distracting (e.g., Horrey and Wickens, 2006; Strayer and Johnston, 2001). As such, restricting the use of hand-held cell phones may not be a sufficient countermeasure for distracted drivers. Drivers are also subject to many different forms of distraction from in-vehicle devices, including mobile "text messaging"—an activity that is likely far more distracting than cell phone conversations. To date, there have been relatively few legislative movements specific to these devices (or activities; Washington, U.S. is one state that has recently passed legislation to ban text messaging).

Given these issues associated with outright bans, alternate means of mitigating driver distraction may be warranted. Unfortunately, in contrast to studies on the magnitude of distraction effects, the examination of potential driver-based strategies for reducing distraction has lagged. These approaches may benefit from an understanding of how a driver perceives distraction as it relates to their actual performance. Perception or awareness of distraction effects may influence drivers' decisions or their willingness to engage in distracting activities while on the road. For example, drivers that are not calibrated with respect to the magnitude of distraction effects may engage in activities because they do not realize their performance is compromised. Put another way, drivers may be overconfident in their ability to drive while distracted. In a recent survey, Wogalter and Mayhorn (2005) found that cell phone users tended to be more optimistic about their ability to deal with distractions than they were about

* Corresponding author. Tel.: +1 508 497 0237; fax: +1 508 435 0482.
E-mail address: william.horrey@libertymutual.com (W.J. Horrey).

0001-4575/$ – see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2007.09.004

other drivers' abilities. Pöysti et al. (2005) found that younger drivers, males drivers, and those that rated themselves as more skilful were more likely (and willing) to use a cell phone in traffic.

In an earlier study, Lesch and Hancock (2004) found that a priori ratings of confidence in dealing with distracting tasks were not related to actual performance while distracted. Moreover, they did not find any relationship between subjective ratings of performance and task demands and actual performance, suggesting that some drivers may not be aware of performance decrements while distracted. In the Lesch and Hancock (2004) study, younger and older drivers performed a memory and recall test while navigating an instrumented vehicle around a closed test track (see also Hancock et al., 2003). During some trials, drivers were distracted by a visual prompt that occurred in conjunction with a changing traffic light. Measures of stopping performance were assessed and compared against confidence ratings. Overall, there was little association between performance and confidence in female drivers; however, the disparities were greatest for older females. The degree to which drivers are calibrated to distraction effects of non-visual distractions remains unknown. As noted above, many studies have demonstrated that cognitive engagement (e.g., conversation) leads to performance loss in the driving task (e.g., Alm and Nilsson, 1994; Strayer and Johnston, 2001).

Finally, understanding how drivers perceive or misperceive distraction may also help inform the application of advanced in-vehicle automation aimed at mitigating distraction (e.g., Donmez et al., 2007). It is likely that the gap between drivers' estimates of distraction effects and their actual performance would play an important role in determining the degree of user trust, reliance and compliance with such systems (Parasuraman and Riley, 1997; Lee and Moray, 1994; Lee and See, 2004).

1.1. Current study

In the current study, we wished to establish how well-calibrated drivers are with respect to distraction effects—that is, whether drivers' estimates of the magnitude of distraction reflect actual performance decrements on several driving tasks. We expand on the results from Lesch and Hancock (2004) using a distraction task which did not require visual processing. Furthermore, we were interested in whether subjective estimates of distraction and actual distraction varied as a function of phone type (hand-held, hands-free). Younger and older drivers in this study drove an instrumented van on a closed test track while performing a continuous mental arithmetic task on a hand-held or hands-free cell phone. Subjective measures of distraction effects were recorded and compared to actual performance on multiple measures of driving performance.

2. Methods

2.1. Drivers

Forty drivers, divided into two age groups, were recruited for this study through advertisements in local newspapers. The younger driver group consisted of 20 drivers between the ages of 18 and 34 (M = 21.8 years, S.D. = 4.3). The older driver group consisted of 20 drivers between the ages of 55 and 82 (M = 64.1 years, S.D. = 7.7). Males and females were balanced across the two age groups. The mean number of years driving experience was 5.5 for the younger drivers and 45.7 for the older drivers. On average, younger drivers drove 14,200 miles (22,850 km) per year while older drivers drove 15,800 miles (25,430 km) annually. All had normal or corrected-to-normal visual acuity. Drivers were paid US$ 20 for each hour of participation and were given a US$ 20 bonus as incentive for good performance on the secondary tasks.

2.2. Materials

The experimental protocol was implemented on a 0.8 km (0.5 mile) closed-loop test track. The two-lane track was delineated to allow for continuous driving. A signalized intersection, located at the end of a straightaway, was controlled through track-based infrared sensors and vehicle-based GPS and speed information transmitted via wireless (DSS) modem.

Pace clocks were used for a longitudinal (speed) control task. Five pace clocks were positioned at various points around the track. These 0.5 m diameter clocks were mounted on 1.8 m towers and placed immediately to the left of the drivers' lane (see Fig. 1). The bottom half of the clock was green and the top half was red. The arrow hand moved around the clock at a constant rate and completed the full rotation every 12 s, on average (ranging from 10 to 14 s, between clocks). The task instructions are described below.

The instrumented vehicle was a 2002 Ford Windstar minivan, outfitted with several sensors and computers. A PC computer rack, mounted directly behind the driver, controlled and coordinated various aspects of the data acquisition and stimulus presentation. Vehicle data were collected at 30 Hz from multiple sources, including the vehicle's front electronic module (FEM), roof-mounted GPS (by Garmin), tri-axial accelerometer (by Crossbow Technologies), and lane tracking cameras (Barickman and Stoltzfus, 1999). Several cameras were positioned to record various interior and exterior viewpoints. A flip down 17 cm LCD touch screen (by Xenarc), mounted in the interior roof console, and a hand-held numeric keypad allowed the

Fig. 1. Pace clocks and instrumented van. Nearest clock currently indicates that driver is permitted to pass.

experimenter to control the computers from the front passenger seat. An extra brake pedal was also installed on the passenger side. Auditory stimuli for the phone tasks were presented via handheld phone (Avaya 3626) or through speakers mounted behind the driver (Sony SRS-T100PC).

2.3. Procedure

At the start of the 2 h session, drivers completed an informed consent form. Color vision and visual acuity were tested using a Titmus Vision Tester (Titmus Optical Inc., Chester, VA). Drivers then completed a brief demographic questionnaire and several scales. In addition to basic demographic information, drivers were asked to rate their confidence in dealing with distractions while driving. Drivers also compared themselves to the average US driver on a number of items pertaining to skill and safety (based on Horswill et al., 2004).

After completing the questionnaires, drivers were introduced to the safety features of the instrumented van and given several minutes of practice to familiarize themselves with the handling of the vehicle, the various driving tasks, and the track layout. For all conditions, drivers were free to select their speed; however, they were instructed not to exceed 48 kph (30 mph). Drivers were further instructed to keep the vehicle positioned in the center of the lane as they navigated the course. For the pace clock (speed control) task, drivers were told to adjust their speed during the approach to a clock, either by accelerating or braking, in order to pass the clock when the arrow indicator was in the green portion of the clock. In other words, they were to avoid passing the clock when the arrow was in the red portion. Drivers were further instructed to avoid bringing the vehicle to a full stop and to avoid exceeding a speed of 48 kph (30 mph) for this task. The traffic light at the signalized intersection changed from green to red on a random subset of trials (37.5%). For these trials, drivers were instructed to bring the vehicle to a complete stop as quickly as they could and to try to stop before they reached a stop line marked by two traffic cones. There was no yellow sequence in the light change in order to discourage drivers from trying to run the light. The timing of the light change, controlled through GPS, track sensors and vehicle speed, varied between 4.5 and 5.5 s before the intersection. Thus, drivers had to respond in a timely manner and brake firmly (with a deceleration of approximately 0.33 g at 40 kph), but they were not required to be overly aggressive in their response (i.e., locking the brakes or skidding).

During the experimental blocks, drivers were asked to perform a concurrent phone task—a variation of the Paced Auditory Serial Addition Task (PASAT; Gronwall, 1977). Similar arithmetic tasks have been used in previous studies on driver distraction (e.g., Brookhuis et al., 1991; Patten et al., 2004). Numbers between 1 and 9 were randomly presented every 7 s, either through the hand-held phone or through the vehicle speakers (hands-free). The task involved adding two consecutive numbers and responding verbally. That is, participants were required to call out the sum of the current number and the number they heard previously. To encourage task completion, drivers were told they would receive a small bonus for each correct response (5 cents per, up to US$ 20 over the course of the study). All drivers received the full bonus at the end of the experiment, regardless of performance.

Drivers completed three blocks of 8 laps, each lasting approximately 15 min. In two blocks, drivers performed the PASAT (once using the hand-held phone, once using the hands-free system). In the other block, drivers performed the driving tasks alone, with no PASAT. Drivers also completed a 3 min baseline block for the PASAT. This block was performed while the vehicle was parked. Drivers were offered a short rest break in between each block. The order of blocks was counterbalanced across driver. There were also two additional blocks (1 driving and 1 non-driving baseline) that are not related to the current discussion or design.

This experiment employed a $3 \times 2 \times 2$ mixed design, with the within-subject variable of Task condition (Baseline, Hand-held PASAT, Hands-free PASAT) and the between-subject variables of Age (young, older) and Gender. Throughout the blocks, measures of lane keeping, pace clock accuracy, and stop light response time and errors were recorded. In between blocks, drivers completed a modified NASA TLX (Hart and Staveland, 1988) that was expanded to include subjective ratings of performance for all of the relevant driving sub-tasks (lane keeping, pace clocks, stopping task) and for the PASAT.

3. Results

There was a small degree of data loss due to occasional equipment or computer failures—a loss that is reflected in the differences in degrees of freedom across the various analyses. In general, data from no more than one driver were missing for a given group across the different measures and conditions. Analyses were conducted with listwise deletion of cases with missing data. An alpha level of .05 was used for all statistical tests.

3.1. Performance decrements

Prior to examining drivers' calibration of distraction effects, we established the presence of dual-task costs associated with the cell phone tasks (i.e., manipulation check). Driving performance was assessed through multiple dependent measures, including variability in lane keeping, accuracy on the pace clock task, brake response time to the changing traffic light, and stop light errors (described below). In general, we used a mixed ANOVA with the variables of Task, Age and Gender (using SPSS Version 11.0.1; see Table 1). Across the various measures, there were no effects of age or gender, or any significant interactions. Mean data are shown in Table 2.

Measures of lane position were sampled at 30 Hz along three straight sections of the track (totaling approximately 345 m per lap). We used a simple moving average to smooth the data (with a window of 30 samples, equivalent to one second of data). The cumulative samples for each block were used to calculate the variability in lane keeping. The mixed ANOVA revealed a main effect for Task (Table 1), with increased variability in the hand-held and hands-free task conditions relative to baseline ($t(37) = 2.7$, $p = .01$; $t(38) = 4.0$, $p < .001$; respectively).

Table 1
ANOVA results for various dependent measures related to driving performance

Source	Brake response time			Pace clock accuracy			Variability in lane keeping[a]		
	d.f.	MSE	*F*	d.f.	MSE	*F*	d.f.	MSE	*F*
Between-subjects									
Age (A)	1,30	0.08	1.5	1,34	296.1	2.4	1,33	1.6E-04	0.3
Gender (G)	1,30	0.03	0.5	1,34	0.7	0.01	1,33	3.2E-04	0.6
A × G	1,30	0.06	1.2	1,34	7.1	0.06	1,33	3.4E-06	0.01
Within-subjects									
Task (T)	2,60	0.06	5.8**	2,68	1261.2	27.1**	2,66	7.9E-04	8.5**
T × A	2,60	0.005	0.5	2,68	10.5	0.2	2,66	1.9E-04	2.0
T × G	2,60	0.001	0.1	2,68	69.9	1.5	2,66	1.9E-04	2.0
T × A × G	2,60	<0.001	0.04	2,68	5.3	0.1	2,66	1.8E-07	0.002

[a] Prior to analysis, variability data was transformed using a log transformation, $Y' = \log_{10}(Y+1)$, to increase normality (based on Kirk, 1982). $^{*}p < .05$. $^{**}p < .01$.

The difference between the two phone types was not statistically significant ($t(36) = 1.6$, $p = .13$). These data are shown in Table 2.

Accuracy in the pace clock task was determined by the percentage of correct clock events per block (out of 45). Correct events were scored when the driver passed the clock when the arrow was in the green portion. Passing the clock in the red portion, exceeding the speed limit during the approach to a clock tower, and coming to a full stop were considered errors on this task (as per the task instructions). As shown in Tables 1 and 2, there was a significant main effect for Task, with drivers making more pace clock errors in the hand-held and hands-free conditions compared to baseline driving ($t(37) = 6.7$, $p < .001$; $t(40) = 6.7$, $p < .001$, respectively). There was a slight advantage for the hands-free over the hand-held phone, however this difference did not reach conventional levels of statistical significance ($t(37) = 1.7$, $p = .09$).

For the stopping task, brake response time (RT) was measured from the onset of the light change until a braking response was detected. As for lane keeping and pace clock accuracy, there was a significant main effect for task condition (Table 1), with slowed response times relative to baseline conditions for the hand-held ($t(34) = 3.7$, $p = .001$) and hands-free conditions ($t(38) = 3.3$, $p = .002$; see Table 2). There were no significant differences between the two phone types ($t(34) = .08$, $p = .41$).

Stop light errors included failures to comply with the light (i.e., running the red light), failures to stop before the stop line, and instances where the driver stopped, then proceeded through the intersection before the light returned to green. Failure to stop before the stop line was the predominant error type (over 90% of the observed errors). The proportion of stopping errors yielded data that were not normally distributed, so we used Wilcoxon signed ranks *t*-tests to examine differences across the task conditions. Compared to the baseline driving condition, there were more stop light errors in both the hand-held ($Z(37) = 3.4$, $p = .001$) and hands-free conditions ($Z(39) = 3.6$, $p < .001$). Again, there were no differences between the two phone conditions ($Z(37) = 0.1$, $p = .93$).

Therefore, across all measures of driving performance we observed decrements due to distraction. In general, and as reported elsewhere, there were no differences between hand-held and hands-free cell phones across our performance measures (Strayer and Johnston, 2001; Horrey and Wickens, 2006). Furthermore, there were no effects of age or gender, or any significant interactions.

Table 2
Driving performance as a function of task condition

Task condition	Brake response time (s)	Pace clock accuracy (%)	Variability in lane keeping[a] (m)	Stop light errors (%)
Baseline	0.88 (0.02)	82.2 (1.2)	0.20 (0.01)	26 (6)
Hand-held	0.96 (0.03)	71.1 (1.6)	0.22 (0.01)	48 (6)
Hands-free	0.94 (0.03)	73.9 (1.2)	0.23 (0.01)	48 (5)

[a] To ease comprehension, these values reflect the raw variability data (vs. log-transformed data). Standard errors appear in parentheses.

3.2. Subjective ratings

We also analyzed the raw estimates of performance on the various driving tasks measured at the end of each block. Subjective performance ratings were entered in a mixed ANOVA for Age, Gender and Task, shown in Table 3. Across all the driving tasks, there were no significant main effects for Age and Gender, or any significant interactions. However, there was a significant main effect of Task on the subjective performance ratings for the stopping, pace clock, and lane keeping tasks.

As shown in Table 4, performance in the driving baseline was rated more favorably than in either distraction condition. That is, performance in the hand-held and hands-free was generally rated lower than single-task driving blocks (pairwise comparisons against baseline, $p < .05$). These data suggest that, overall, drivers *do* consider the distracting effects of in-vehicle activities in their performance estimates. Ratings across phone type were equitable on all measures except for the stopping task, where performance with the hand-held phone was rated as more successful than the hands-free condition ($t(36) = 2.2$, $p < .05$). Note that there were no observed differences between these two conditions based on the performance data described above.

Although the overall aggregate ratings reflected degraded performance across the different task conditions, what is of greater interest is whether an individual driver's assessment of distraction reflects their actual performance on a given task. We examine this issue in the following section.

Table 3
ANOVA results for the subjective performance ratings for the various driving tasks

Source	Stopping task			Pace clock task			Lane keeping task		
	d.f.	MSE	*F*	d.f.	MSE	*F*	d.f.	MSE	*F*
Between-subjects									
Age (A)	1,33	1301.5	1.1	1,33	50.2	0.06	1,33	153.6	0.2
Gender (G)	1,33	674.0	0.6	1,33	384.9	0.4	1,33	135.9	0.1
A × G	1,33	1548.5	1.3	1,33	209.9	0.2	1,33	44.9	0.05
Within-subjects									
Task (T)	2,66	1438.0	19.2^{**}	2,66	2271.8	16.3^{**}	2,66	938	7.5^{**}
T × A	2,66	46.3	0.6	2,66	69.3	0.5	2,66	63.3	0.5
T × G	2,66	184.6	2.5	2,66	22.3	0.2	2,66	153.3	1.2
T × A × G	2,66	95.7	1.3	2,66	29.3	0.2	2,66	4.9	0.04

Note: $^{*}p < .05$; $^{**}p < .01$.

3.3. Calibration to distraction effects

To examine the calibration of drivers with respect to the distraction effects, we first calculated estimated and actual distraction effects from the subjective performance ratings and from observed performance along the various measures of driving performance. Subjective estimates of distraction were expressed as a % difference between the self-rated performance on a given task (e.g., lane keeping) in the driving baseline block and self-rated performance in the distraction block. We used the baseline condition as a reference point in order to control for any systematic biases in ratings. Likewise, actual distraction effects were expressed as a % difference in performance on a given task from the baseline block to the dual-task block. Thus, negative values denoted an actual or estimated loss in performance for the distraction condition relative to baseline, whereas a positive value indicated a gain in actual or estimated performance.

Next, we examined the relationship between subjective estimates of distraction effects and actual distraction effects. If drivers are well-calibrated, we would expect a positive correlation between the estimated and actual distraction effects. That is, larger estimations of distraction effects would correspond to larger observed performance decrements. However, if drivers are not well-calibrated to the magnitude of distraction effects, then there would be no such relationship or a negative relationship may exist.

The results from a correlational analysis are shown in Table 5. Overall, there were no significant relationships between estimates of distraction effects and actual performance decrements, lending support to the notion that drivers were not well-calibrated to the distracting effects of a concurrent in-vehicle task. There was one significant relationship for stopping errors in the hands-free condition. Importantly, however, this relationship is in the negative direction, again underlying poor calibration. That is, drivers that estimated the smallest performance decrements were actually exhibiting the largest ones! An examination of the 95% confidence intervals around the correlations in Table 5 suggests that, in general, the poor calibration of drivers to the effects of distraction were consistent across phone type and dependent measure (Note: although there were no differences in calibration across measure, there may have been differences in *resolution*—the degree to which drivers can distinguish between correct and incorrect judgments (see Murphy, 1973)—given differences in the available feedback).

A breakdown of these relationships by driver group revealed some interesting findings. For example, older male drivers were actually well-calibrated to the magnitude of distraction effects for the stopping task. In contrast, younger males showed some significant associations in the opposite direction. That is, young male drivers that thought they were doing better were actually doing worse than others. In general, female drivers did not exhibit any significant relationships between estimated and actual performance loss (as shown by Lesch and Hancock, 2004).

Table 4
Subjective performance ratings (%) as a function of task condition

Condition	Stopping task	Pace clock task	Lane keeping task
Baseline	78.6 (3.5)	65.0 (3.6)	70.6 (3.6)
Hand-held	71.2 (3.9)	51.8 (3.1)	64.5 (3.5)
Hands-free	66.2 (3.0)	51.0 (3.0)	60.5 (2.7)

Note: Standard errors appear in parentheses.

3.4. Under-estimators and over-estimators

Given the trends for drivers not to be well-calibrated with respect to the distraction effects, we conducted an exploratory analysis to determine whether individuals who tended to underestimate the effects of distraction differed on a number of measures from individuals who overestimated the effects of distraction. Underestimating the impact of distraction on performance could have dire consequences in the traffic setting. For this analysis, we calculated a difference score by subtracting the actual distraction effects (in %) from the estimated distraction effects (see Lichtenstein and Fischhoff, 1977; Baranski and Petrusic, 1994, 1995). The score was averaged across phone type and dependent measure in order to simplify the interpretation of the subsequent analysis and to identify those individuals who exhibited a general tendency towards over- or under-estimation. Individuals who scored less than −5% were included in the under-estimator group ($N = 15$) – the magnitude of actual dis-

Table 5
Correlations between the magnitude of estimated and actual distraction effects

Driver group	Δ Brake response time		Δ Pace clock accuracy		Δ Lane keeping		Δ Stop light errors	
	Held	Free	Held	Free	Held	Free	Held	Free
Overall	0.06 (−0.28, 0.39)	−0.02 (−0.33, 0.31)	0.16 (−0.18, 0.46)	0.10 (−0.22, 0.40)	0.25 (−0.09, 0.53)	0.23 (−0.10, 0.51)	0.11 (−0.23, 0.42)	−0.38* (−0.62, −0.07)
Young males	−0.65 (−0.93, 0.10)	−0.33 (−0.79, 0.38)	−0.07 (−0.70, 0.62)	0.30 (−0.41, 0.78)	0.11 (−0.56, 0.69)	0.17 (−0.52, 0.72)	0.05 (−0.64, 0.69)	−0.91** (−0.98, −0.66)
Older males	0.82** (0.34, 0.96)	0.45 (−0.26, 0.84)	0.40 (−0.36, 0.84)	−0.22 (−0.74, 0.48)	0.62 (−0.08, 0.91)	0.40 (−0.31, 0.82)	0.30 (−0.46, 0.80)	0.71* (0.15, 0.93)
Young females	0.07 (−0.62, 0.70)	0.13 (−0.58, 0.73)	0.42 (−0.34, 0.85)	0.51 (−0.18, 0.86)	0.26 (−0.44, 0.77)	−0.56 (−0.89, 0.17)	0.34 (−0.37, 0.80)	−0.61 (−0.91, 0.09)
Older females	0.38 (−0.38, 0.83)	−0.04 (−0.65, 0.60)	−0.23 (−0.77, 0.52)	−0.29 (−0.78, 0.41)	0.31 (−0.57, 0.81)	0.53 (−0.25, 0.84)	−0.29 (−0.80, 0.46)	−0.56 (−0.88, 0.11)

Note: The parentheses include the 95% confidence interval (CI). Held = hand-held; free = hands-free.
* $p < .05$.
** $p < .01$.

Table 6
Results of the forward stepwise logistic regression

Variables	*B*	S.E.	Wald	*p*
Constant	1.3	0.70	3.4	.07
(1) Distractibility	−0.05	0.02	4.0	.05

Excluded variables	Score	*p*
Age	1.1	.30
Gender	1.7	.19
(2) Safety	2.4	.12
(3) Skill	0.001	.97
(4) Ability to Cope	0.3	.60
(5) Confidence	1.8	.18
(6) Hand-held Safety	0.1	.73
(7) Hands-free safety	1.1	.30
(8) Hand-held Ease	0.1	.72
(9) Hands-free Ease	0.01	.91

Note: All responses were made along a continuous scale. (1) "How distractible are you when driving?", (2) "How safe are you?", (3) "How skillful are you?", (4) "How good are you at dealing with distraction while driving?", (5) "How confident are you in dealing with distracting tasks while driving?", (6 and 7) "How safe is using a [hand-held/hands-free] cellular phone while driving?", (8 and 9) "How easy is using a [hand-held/hands-free] cellular phone while driving?" For Questions 1–4, drivers compared themselves to the average U.S. driver.

traction effects were greater than their estimates. Individuals who scored more than +5% were assigned to the over-estimator group ($N = 16$) – the magnitude of the distraction effects were less than what was estimated. As noted above, the latter is more desirable given the safety implications for failing to appreciate the consequences of distraction. We did not include those individuals whose difference score fell within 5% of zero ($N = 9$), focusing rather on the more extreme over- and under-estimators.

Next, we examined whether there were group differences in driving performance and subjective ratings in the distraction conditions. Independent samples *t*-tests did not reveal any group differences in braking RT ($t(29) = 0.7$, $p = .50$), pace clock accuracy ($t(29) = 0.2$, $p = .83$), and stop light errors ($t(29) = 0.3$, $p = .80$), indicating that over- and under-estimators had equitable performance on the current driving tasks. However, there were marginally significant differences in the subjective ratings of performance on the stopping and pace clock tasks ($t(29) = 2.0$, $p = .06$; $t(29) = 2.0$, $p = .06$, respectively), with under-estimators rating their own performance as higher than over-estimators. This pattern of results suggests that differences between the under- and over-estimators are, in fact, due to differences in judgments and not a result of differences in performance.

We then examined whether over- or under-estimators could be predicted from a number of questionnaire items collected from participants, using a forward stepwise logistic regression model. Age, gender, and driver responses to a number of questions pertaining to distraction and driving were included as predictor variables (using a $p < .05$ threshold).

As shown in Table 6, only one variable was entered into the model: drivers' self reported assessment of distractibility. Drivers who tended to underestimate the effects of distraction also rated themselves as more prone to distraction while driving, compared to the average US driver. Interestingly, measures of self-confidence in dealing with distraction (confidence in dealing with distraction; ability to cope with distracting tasks) and driving skill and safety were not significant predictors of under- and over-estimators, nor were initial ratings of the safety and ease of use of hand-held and hands-free cell phones (see Table 6). Thus, the drivers that underestimated the distraction effects did not appear to be more overconfident than over-estimators based on a priori estimates of confidence and skills. Furthermore, age and gender were not significant predictors.

4. Discussion

The main goal of this study was to examine the extent to which drivers were calibrated to the effects of distraction on driving performance. Previous work has reliably demonstrated the adverse impact of concurrent in-vehicle tasks on driving performance, typically showing slowed response times or missed traffic events in the presence of a distracting task (e.g., Alm and Nilsson, 1994; Hancock et al., 2003; Horrey and Wickens, 2006; McKnight and McKnight, 1993; Strayer and Johnston, 2001). However, few have examined the extent to which drivers are aware of these decrements. Understanding drivers' perception and awareness of distraction effects may help inform driver-based or technology-based interventions aimed at mitigating distraction by helping drivers better manage the engagement and disengagement of in-vehicle activities. In contrast, legislative approaches tend to rely more on the presence and magnitude of a given form of distraction with the aim of eliminating the source of distraction altogether (e.g., hand-held cell phones in some areas). Lesch and Hancock (2004) suggest that drivers may not, in fact, be very cognizant of distraction effects. The current work sought to expand on this earlier study using a cognitive task and multiple phone configurations.

In the current study, younger and older drivers were asked to complete a hand-held or hands-free cell phone task while navigating a closed test-track in an instrumented vehicle. Compared to baseline driving, we observed dual-task decrements on all measures of driving performance, replicating previous results (Alm and Nilsson, 1994; McKnight and McKnight, 1993). While normal aging is often associated with degraded performance on many psychomotor tasks, such as response time (e.g., Salthouse, 1996); we did not observe any age-related effects in the current study—an outcome that could be related to the fact that our age groups included a wide range of ages (i.e., had a higher degree of heterogeneity leading to decreased statistical power associated with an age effect). It is possible that the adults in our sample were healthier and more active than average. Furthermore, studies have shown that older adults, in some cases, are able to offset age-related declines through increased driving experience and improved skills (Kramer et al., 2007).

To examine drivers' calibration to distraction effects, we compared drivers' subjective estimates of distraction with actual distraction effects, based on performance decrements along a number of measures of driving performance. The results from our study suggest that, for the most part, drivers are not well-calibrated to the distracting effects of a hand-held or hands-free cell phone conversation, although, the omnibus ANOVAs revealed that drivers rated their performance while distracted as being poorer than without distraction. That being said, across all measures of performance, subjective estimates of distraction effects (derived from ratings of performance across distraction and baseline blocks) were not related to the actual magnitude of distraction (based on observed performance decrements). And, in some cases, the subjective measure of distraction was in the opposite direction of the actual distraction effect. That is, drivers that estimated the smallest (or no) distraction effects exhibited the largest ones. In general, a disconnect between performance and awareness was consistent across driving measure and phone type.

Analysis of the driver groups, though limited by a small sample size, showed some differences. For female drivers, there were no significant relationships across all measures and phone types (Lesch and Hancock, 2004); however, there were some mixed results for males. For example, on some measures young and older males revealed a nearly equal, but opposite, relationship between estimated and actual distraction effects. Younger males were poorly calibrated, with the worst performers rating themselves as exhibiting the smallest decrements. In contrast, older males were fairly well-calibrated to the distraction effects. It follows that younger male drivers may be an important group for targeted remediation.

Given the safety implications for those individuals who underestimate the effects of distraction on their performance, we conducted an exploratory analysis of these sub-groups. These results suggest that drivers who tend to overestimate distraction effects did not differ from those that underestimated the effects of distraction on a number of measures related to confidence and perceived skills. For example, there were no differences in self-rated confidence in dealing with distracting tasks, their perceived driving skills, and their perceptions of the challenge associated with cell phone use while driving. Thus, overconfidence does not appear to underscore errors in calibration to distraction effects—at least using the current metrics. Rather, the mechanism for failures in calibration may be due to a lack of awareness in performance or failure of perception, as opposed to more general biases related to self-confidence. This may stem from a reduction in the available mental resources under dual-task conditions that would normally be used in support of situation awareness (Wickens, 2001).

4.1. Implications

The willingness to engage in distracting activities may be a function of drivers' perception of performance decrements. Drivers may engage in distracting activities simply because they do not realize that their performance is degraded or they may be overconfident in their skills and their ability to deal with distractions while behind the wheel (Wogalter and Mayhorn, 2005). The results from the current study and from previous work by Lesch and Hancock (2004) suggest that drivers may be poorly calibrated to distraction effects.

Strategies for improving drivers' calibration to distraction effects include both driver- and technology-based approaches. For the former, training to recognize or attend more closely to their driving activities may help drivers' determine when their performance is below “baseline”. For example, Lichtenstein and Fischhoff (1980) found that intensive performance feedback could help improve calibration in people who were initially overconfident in their judgments (cf. Baranski and Petrusic, 1994). Some technological innovations are intended to monitor driver performance and behaviors in real time and provide alerts when the system infers that a distracting activity is inappropriate (e.g., Donmez et al., 2007). Thus, these systems have some authority in deciding when the distraction effects of in-vehicle activities are too severe. However, discrepancies between the driver and the system’s estimated level of distraction could create some conflicts. For example, a driver’s mental model of the system may be violated if warnings are issued for events for which a driver does not believe to be problematic. The consequences of such a mismatch could include system disuse deriving from reduced trust in the system (Lee and See, 2004; Lee and Moray, 1994; Parasuraman and Riley, 1997) or worse (Sarter and Woods, 1995). Providing the raw data underlying the system functions may be a key consideration (cf. Oskamp, 1965).

One question that remains is whether drivers would continue to perform in-vehicle tasks if they were perfectly calibrated to the distraction effects. In other words, the issue could become one of risk tolerance and the degree to which drivers are willing to take on the additional risk of the in-vehicle task. We also do not know how well-calibrated drivers are for distracting tasks that may be more practiced. For example, we used a mental arithmetic task, which would be novel for our participants (versus a very familiar task). Finally, different in-vehicle task properties or characteristics may impact drivers’ calibration differentially. For example, Baranski and Petrusic (1994) found that confidence varied as a function of task difficulty, with observers showing overconfidence when then task was difficult and underconfidence when it was easy (see also Soll, 1996). Identifying those factors that contribute to improved calibration is an area for further examination.

Acknowledgements

The authors are grateful to Peter Teare, Eric Jones, Richard Holihan, and Ed Correa for their maintenance and upkeep of the instrumented van and other equipment. We also thank Raymond McGorry and William Shaw for their comments and suggestions regarding an earlier version of this paper. A subset of data from the current manuscript was presented at the 4th International Driving Symposium on Human Factors in Driving Assessment, Training, and Vehicle Design.

References

Alm, H., Nilsson, L., 1994. Changes in driver behaviour as a function of handsfree mobile phones—A simulator study. Accid. Anal. Prev. 26, 441–451.

Baranski, J.V., Petrusic, W.M., 1994. The calibration and resolution of confidence in perceptual judgments. Percept. Psychophys. 55 (4), 412–428.

Baranski, J.V., Petrusic, W.M., 1995. On the calibration of knowledge and perception. Can. J. Exp. Psychol. 49 (3), 397–407.

Barickman, F.S., Stoltzfus, D.L., 1999. A Simple CCD-Based Lane Tracking System (Report No.1999-01-1302). Society of Automotive Engineers, Warrendale, PA.

Brookhuis, K.A., de Vries, G., de Waard, D., 1991. The effects of mobile telephoning on driving performance. Accid. Anal. Prev. 23, 309–316.

Donmez, B., Boyle, L.N., Lee, J.D., 2007. Safety implications of providing real-time feedback to distracted drivers. Accid. Anal. Prev. 39 (3), 581–590.

Gronwall, D.M.A., 1977. Paced auditory serial-addition task: A measure of recovery from concussion. Percept. Mot. Skills 44, 367–373.

Hancock, P.A., Lesch, M., Simmons, L., 2003. The distraction effects of phone use during a crucial driving maneuver. Accid. Anal. Prev. 35 (4), 501–514.

Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Amsterdam, North-Holland, pp. 139–183.

Horrey, W.J., Wickens, C.D., 2006. Examining the impact of cell phone conversations on driving using meta-analytic techniques. Hum. Factors 48 (1), 196–205.

Horswill, M.S., Waylen, A.E., Tofield, M.I., 2004. Drivers’ ratings of different components of their own driving skill: A greater illusion of superiority for skills that relate to accident involvement. J. Appl. Soc. Psychol. 34 (1), 177–195.

Kirk, R.E., 1982. Experimental Design: Procedures for the Behavioral Sciences. Wadsworth, Belmont, CA.

Kramer, A.F., Cassavaugh, N., Horrey, W.J., Becic, E., Mayhugh, J., 2007. Influence of age and proximity warning devices on collision avoidance in simulated driving. Hum. Factors 49 (5), 935–949.

Lee, J.D., Moray, N., 1994. Trust, self-confidence, and operators’ adaptation to automation. Int. J. Hum. Comput. Studies 40, 153–184.

Lee, J.D., See, K.A., 2004. Trust in automation: Designing for appropriate reliance. Hum. Factors 46 (1), 50–80.

Lesch, M.F., Hancock, P.A., 2004. Driving performance during concurrent cell-phone use: Are drivers aware of their performance decrements? Accid. Anal. Prev. 36, 471–480.

Lichtenstein, S., Fischhoff, B., 1977. Do those who know more also know more about how much they know? The calibration of probability judgments. Org. Behav. Hum. Perform. 20, 159–183.

Lichtenstein, S., Fischhoff, B., 1980. Training for calibration. Org. Behav. Hum. Perform. 26, 149–171.

McKnight, A.J., McKnight, A.S., 1993. The effect of cellular phone use upon driver attention. Accid. Anal. Prev. 25, 259–265.

Murphy, A.H., 1973. A new vector partition of the probability score. J. Appl. Meteorol. 12, 595–600.

Oskamp, S., 1965. Overconfidence in case-study judgments. J. Consult. Psychol. 29, 261–265.

Parasuraman, R., Riley, V., 1997. Humans and automation: Use, misuse, disuse, abuse. Hum. Factors 39 (2), 230–253.

Patten, C.J.D., Kircher, A., Östlund, J., Nilsson, L., 2004. Using mobile telephones: Cognitive workload and attention resource allocation. Accid. Anal. Prev. 36, 341–350.

Pöysti, L., Rajalin, S., Summala, H., 2005. Factors influencing the use of cellular (mobile) phone during driving and hazards while using it. Accid. Anal. Prev. 37, 47–51.

Salthouse, T.A., 1996. Processing-speed theory of adult age differences in cognition. Psychol. Rev. 103, 403–428.

Sarter, N.B., Woods, D.D., 1995. How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Hum. Factors 37 (1), 5–19.

Soll, J.B., 1996. Determinants of overconfidence and miscalibration: The roles of random error and ecological structure. Org. Behav. Hum. Decis. Processes 65 (2), 117–137.

Strayer, D.L., Johnston, W.A., 2001. Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular phone. Psychol. Sci. 12 (6), 462–466.

Sundeen, M., 2005. Cell Phones and Highway Safety: 2005 State Legislative Update. National Conference of State Legislatures, Denver, CO.

Wickens, C.D., 2001. Workload and situation awareness. In: Hancock, P., Desmond, P. (Eds.), Stress, Workload, and Fatigue. Lawrence Erlbaum, Mahwah, NJ, pp. 443–454.

Wogalter, M.S., Mayhorn, C.B., 2005. Perceptions of driver distraction by cellular phone users and nonusers. Hum. Factors 47 (2), 455–467.