

PERFORMANCE DEGRADATION DUE TO AUTOMATION IN TEXTING WHILE DRIVING

B. D. Sawyer and P.A. Hancock,
University of Central Florida, Department of Psychology
Orlando, Florida, USA
Email: sawyer@inhumanfactors.com

Summary: Previous research concerning the use of cell phones indicates that physical manipulation of the phone in conjunction with the cognitive need to compose a message together contribute to driving performance degradation. We have suggested that automated assistive text entry schemes such as Nokia's T9 may mitigate some of these identified costs. In this work drivers in a simulator drove and texted using either the assistive T9 system or an unassisted multitap system. Contrary to previous pilot findings participants showed greater degradation of driving performance when using the automated assistive T9 than the unassisted multitap. Findings support the idea that cognitive composition of a message combined with entry interface automation contributed to driving performance degradation. It further implies that the costs of that automation may exceed the benefits.

INTRODUCTION

A growing number of studies have demonstrated a driving detriment due to text messaging (Sawyer & Hancock, 2012, Sawyer, 2010). Such performance diminution involves changes in speed, steering and lane position variability, estimates of ongoing Time to Collision (TTC) and a variety of other measures of driver behavior (Crisler, et al, 2008; Hosking, Young, & Regan, 2009; Tijerina et al, 1995). However, as is true in vocal cell phone use, the presence of the phone in hand of an individual introduces a relatively minimal level of load when compared to the substantial contributions of factors like the cognitive load of language, its conception and delivery (Strayer & Johnston, 2003). Indeed, Sawyer's (2010) evidence suggest that in text messaging the source of the load could likewise be cognitive language processing.

In text messaging, a cell phone user must compose their message, and then access the phone's interface to convert that message into text. Therefore, although the presence of the phone in hand may not constitute any significant source of cognitive load, nonetheless, the interface that allows that individual to convey to the phone software the composed message seems likely to contribute cognitive load to the driving process. Given that the manual text transcription process includes visual and tactile perception, processing and action tasks, tapping the very pools of cognitive resources that driving draws from (see Wickens, 1984), it seems likely that this component of text messaging taxes mental workload significantly. Although this may have no immediate effect upon the driving task, this increase in workload effectively limits the sustainable stable load level (Hancock & Warm, 1989). A spike in overall driver workload, unexpected or not, is therefore more likely to lead to dynamic instability and a failure, and possibly a catastrophic one, in the driving task. Although existing research suggests minimal detriment from manual manipulation driven detriment in text messaging, it is supportable to assert that existing inquiries have looked

at manual manipulations in isolation, without adding the language driven component of composition.

Certainly, many commercial input schemes for cell phones purport to facilitate text entry. Nine key phones originally used a simple scheme named multitap, in which the button is pressed to cycle through letters. Nuance Communications' T9 scheme required only one tap per a letter, and can be considered an assistive technology that, according to Nuance, "...aids in and speeds message composition." (Nokia, 2007) Indeed, typing the word "welcome" on a multitap system requires the user to press 15 keys, while on a T9 system they need press only 7. Given what Miller (1956) and subsequently Baddeley (1975) have reported concerning short term memory, it seems likely that the distinction between multitap and T9 should be important. Further, T9 often allows the user to skip the transliteration that multitap entails, in which the user must code each letter not only to a physical location but to a number of taps.

If manual manipulation in conjunction with language composition contributes to driving degradation, drivers sending text messages might enjoy reduced detriment if the text messaging task were assisted through the predictive T9 system. Indeed, Sawyer and Hancock (2012) in a brief exploratory study found evidence for interaction between manual manipulation, language preparation and delivery as contributing to driver detriment during text messaging. This data also suggested that assistive technologies might mitigate detriment, at least to a degree.

The current inquiry sought to extend the initial pilot work of Sawyer & Hancock (2012). We hypothesized that participants using the predictive T9 technology would exhibit fewer costs to driving while texting. We further hypothesized that a main effect of condition (baseline vs text messaging) would be seen such of it participants in the text messaging condition would exhibit greater costs to the driving task.

METHOD

Experimental Participants

Undergraduate psychology students from UCF were recruited and participated in return for partial class credit. The 60 participants ranged in age from 18 to 34, and were required to have drivers licenses, one year of previous driving experience, and corrected to 20/20 or better vision. A total of 13 participants were dropped from the study, 3 for failing to meet a texting speed criterion, 8 due to equipment failures during experimentation including failure of data collection, and 2 were dropped due to simulation sickness. In all results were adduced from the 47 participants who remained in the study.

Equipment and Training

Participants drove in a fixed platform high fidelity PatrolSim II driving simulator equipped with data collection through Labview. We employed a variant of the pace car cell phone study methodology, a utilized in many studies of cell phone use while driving (Strayer, 2003). Participants were positioned on a simulated roadway with curves and intermittent oncoming

traffic and asked to follow a lead car without passing it. This lead car would brake suddenly at four unexpected intervals over the course of the drive.

Text messages were delivered by a Nokia phone connected to Nokia Communications Suite, and automated by a custom macro. Due to the diversity of cell phones and text input schemes already used by our student sample, as well as questions about previous experience, all participants were trained for two 15 minute periods on the Nokia 8801 cell phone used in the study on the text input scheme they had been randomly assigned to; half to T9 and half to multi-tap. Following training, each was tested to a speed criterion on two panagrams, sentences containing every letter of the alphabet (e.g. Pack my box with five dozen liquor jugs). The criterion of 35 characters per minute (CPM) matched the previous study (Sawyer & Hancock, 2012). Three participants were cut from the study for failing to achieve this speed; notably all three were cut from the T9 Dictionary condition. In response to this pattern, a brief analysis of CPM texting speed by dictionary type (T9 or Multitap) was run. No significant effects on texting speed was found ($F(1, 47) = .498, p = 0.484$), so this appears to be a coincidence. (This outcome, arguably, contradicts Nokia’s claims of enhanced texting speed through T9.) Of the remaining participants, most well exceeded the cut score: texting speeds in excess of 80 CPM were recorded.

Design

The study was a mixed design with 2 (baseline drive vs text messaging) as a participant manipulation x 2 (Multitap vs T9) as a between participants factor.

Participants were entered into either the **T9** or **Multitap** input scheme groups, whereupon the research phone’s preferences were set they were trained accordingly. All participants drove in a **text messaging** condition using their assigned input scheme. The computer controlled texting program sent a series of beginnings to common sayings or nursery rhymes and instructed to reply with the remainder. For example, a participant might be texted “Mary had” and would be expected to respond with “a Little Lamb”. If the participant was unable to recall or did not know the correct ending, they were instructed to text back “I don’t know”. These stimuli were those used in previously (Sawyer & Hancock, 2012) and were selected on the basis that pilot participants were able to finish them more than 90% of the time. The longest outgoing phrase was 19 characters, and the shortest was 11. The longest participant reply was 19 and the shortest was 9. All participants received the same text messages in the same order. The majority were able to respond with the appropriate reply. Text messages were sent throughout the texting condition drive without pause; upon sending the reply the driver would immediately receive another text message. In the **baseline** condition, individuals drove with no phone.

Performance metrics

Four measures of driving were recorded. The first was **Hybrid Response Time (HRT)**, measured as the time from each pace car brake event until a recorded response from the participant, as outlined in Fig. 1.

Minimum Time to Collision (TTC_{min}), a

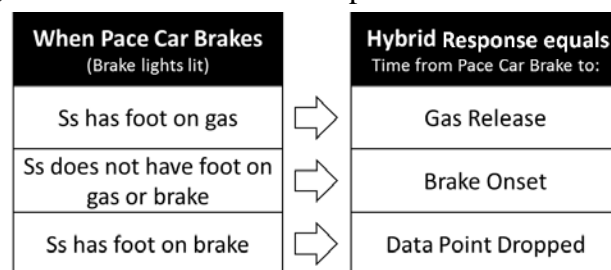


Figure 1. Hybrid Response Time is measured from each Pace Car Brake Event until an indication of reaction, which varies by Ss initial

measure of each driver’s closest approach to the pace car, was windowed over the duration of braking maneuvers, a time stretching from the lead car’s brake event until the nadir of the participant’s speed. The remaining metrics were calculated over a window representing the active part of the drive, calculated from 5 seconds before the first pace car brake event until 1.5 seconds after returning to acceleration following the fourth lead car brake event. **Average Speed** was calculated within this window, as was the **Root Mean Square (RMS) of Steering Angle**, a measure of variance in steering.

RESULTS

A mixed between-within participant ANOVA was conducted to assess the impact of two methods of texting (T9 vs multitap) on driving performance, across four measures (HRT, steering variability, mean speed, and TTC_{min}), as compared to a baseline drive. A significant interaction was found between driving condition (texting vs baseline) and method of texting (T9 vs multitap). A main effect for condition (texting vs baseline) was also found, ($F(4, 42) = 5.76, p = 0.01, partial\ eta\ squared = 0.354$) under which all metrics were significant: HRT ($F(1, 45) = 12.14, p = 0.01, partial\ eta\ squared = 0.21$), RMS of Steering ($F(1, 45) = 4.40, p = 0.03, partial\ eta\ squared = 0.10$), Inverse TTC, ($F(1, 45) = 5.946, p = 0.02, partial\ eta\ squared = 0.117$) and Average Speed ($F(1, 45) = 4.95, p = 0.03, partial\ eta\ squared = 0.10$). Multitap had a moderate effect on driving performance metrics, while the use of T9 had a much stronger effect ($F(4, 42) = 2.972, p = 0.008, partial\ eta\ squared = 0.275$). This pattern held true for HRT, ($F(1, 45) = 6.01, p = 0.02, partial\ eta\ squared = 0.12$) and RMS of Steering, ($F(1, 45) = 5.52, p = 0.02, partial\ eta\ squared = 0.109$). There was a strong but formally insignificant effect on TTC_{min} , ($F(1, 45) = 3.65, p = 0.06, partial\ eta\ squared = 0.08$) but was not significant for Average Speed ($F(1, 45) = 1.42, p = 0.24, partial\ eta\ squared = 0.052$).

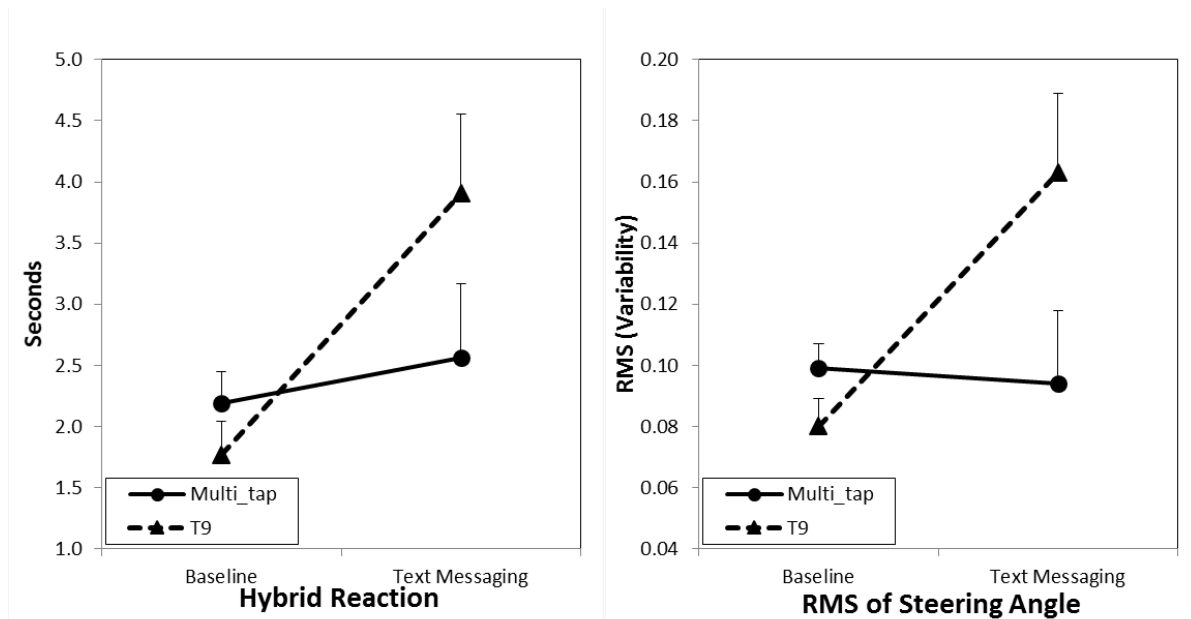


Figure 2. Results showing a main effect of condition (Baseline vs Text Messaging and significant interactions between these conditions and Input Scheme (Multitap vs T9). The error bars represent standard deviation

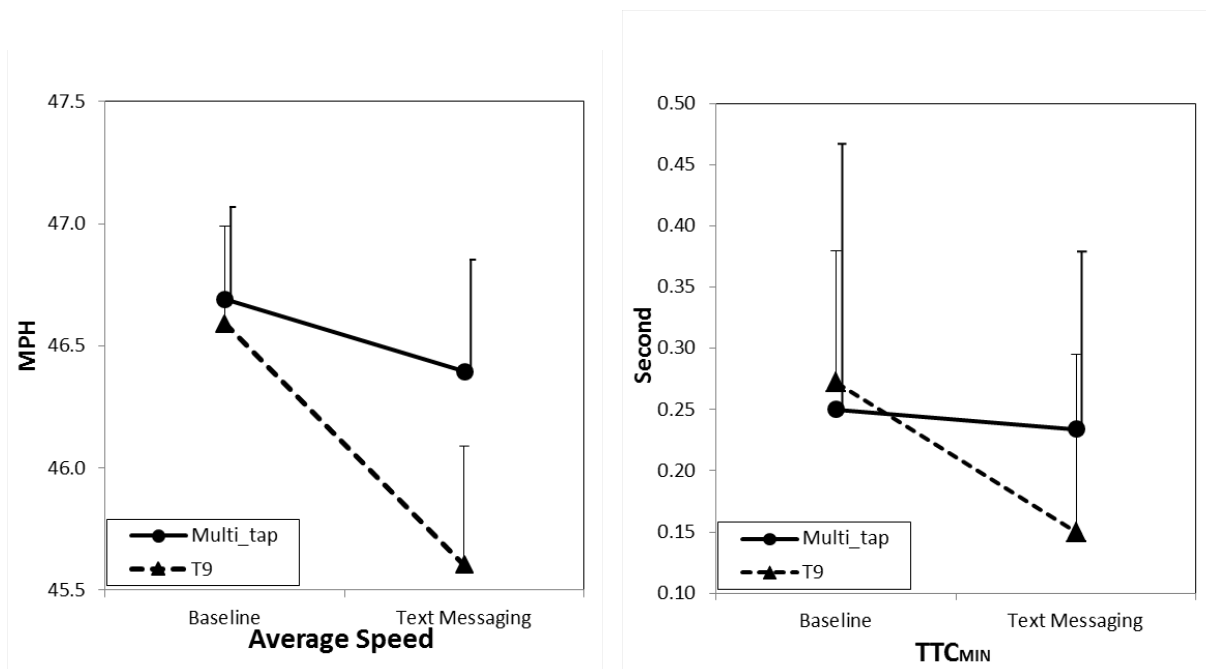


Figure 3. Results not significant under the interaction, but significant under the main effect of condition (Baseline vs Text Messaging). The error bars represent standard deviation

DISCUSSION

Previous findings supporting driving detriment due to text messaging are unequivocally upheld by these data. Our hypothesis that participants using the predictive T9 technology would exhibit less detriment in driving while texting was not supported by the current results. Instead, essentially the opposite pattern appeared. The drivers showed greater driving detriment when using the predictive T9 technology in both HRT and the RMS of steering angle. Although not significant, a similar trend was seen in both average speed and TTC_{min}. These data suggest overall cost to drivers using the assistive T9 technology.

This finding is however not without precedent. Parasuraman and colleagues (1996, for a broad discussion see Parasuraman & Wickens, 2008) for some time espoused the view that the demands of monitoring an automated system can eclipse the benefits to mental workload provided by the automation in the first place. In fact, previous work looking at automation in driving environments has likewise found complex cost-benefit patterns to the introduction of automation. Desmond, Hancock, and Monette (1998) found better recovery from wind gusts when participants had full manual control of the vehicle as compared to an automation assisted drive. Automation in aviation contains numerous examples of these trade-offs (see especially Warm, Dember & Hancock, 1996; Sheridan, 2002). It is therefore possible that whatever advantages T9 bestows are more than offset by the costs of monitoring it in the dual task condition.

Since nine key cell phones are an aging technology, this effect should be evaluated for newer touchscreen devices. Smart phones employ a variety of novel input schemes, and “ease-of-use” and “speed” are some of the most common advertised benefits. Future work needs to evaluate the current effect in these and other in-vehicle devices. It may be, for example, that certain styles of

input automation are more conducive to dual tasking than others, as was shown by Kaber and others (2005) with different styles of automation to help air traffic controllers.

It is also worth discussing whether the manipulations used were themselves suitable proxies for naturalistic texting while driving. The mandate to answer each SMS came in this experiment from the experimenter, and not from any intrinsic desire to see the contents of each text message. Likewise, from an experimental control point of view, it would be superior to window driving data not only around braking events, but around text messaging events themselves. Both these points might be addressed in future work through more flexible windowing and tracking of the participants. Both might shed light on the behaviors and microbehaviors that guide critical decisions to engage in the dangerous behavior of text messaging.

Ultimately, the goal should be not only to understand the danger that text messaging while driving creates, but understand ways in which it might be mitigated. Although legislation may provide an effective deterrent it is likely text communication during crucial primary tasks will remain a feature of human-machine interaction for some time to come. Outside of civilian transportation it may be relevant to commercial or military operations where the added risk can be justified by the context of the work at hand. As such, research should continue to seek out avenues for reducing in-vehicle distraction and its tragic consequences.

REFERENCES

- Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14(6), 575-589.
- Crisler, M. C., Brooks, J. O., Ogle, J. H., Guirl, C. D., Alluri, P., & Dixon, K. K. (2008). Effect of Wireless Communication and Entertainment Devices on Simulated Driving Performance. *Transportation Research Record*, 2069(-1), 48-54.
- Desmond, P., Hancock, P.A., & Monette, J. (1998). Fatigue and automation-induced impairments in simulated driving performance. *Transportation Research Record*, 1628, 8-14.
- Hosking, S. G., Young, K. L., & Regan, M. A. (2009). The effects of text messaging on young drivers. *Human Factors*, 51(4), 582-592.
- Kaber, D. B., Wright, M. C., Prinzel, L. J., & Clamann, M. P. (2005). Adaptive automation of human-machine system information-processing functions. *Human Factors*, 47(4), 730-741.
- Parasuraman, R. (1996). Monitoring of automated systems. *Automation and human performance*.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2, 141-161.
- Tijerina, L., Kiger, S. M., Rockwell, T. H., & Tornow, C. (1995). Workload assessment of in-cab text message system and cellular phone use by heavy vehicle drivers on the road. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, 39, 1117-1121.
- Sawyer, B. (2010). Impact of components of text messaging on simulated driving performance. Honors Thesis, Colorado State University.
- Sheridan, T. B. (2002). *Humans and automation*. New York:Wiley.

- 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(1), 23-32.
- Wickens, C. (1984). Processing resources in attention. In: R. Parasuraman & Davies, Eds., *Varieties of Attention* (pp 63-101). New York: Academic Press.
- Hancock, P.A., & Parasuraman, R. (1992). Human factors and safety in the design of intelligent vehicle-highway systems. *Journal of Safety Research*, 23, 181-198.
- Nuance Communications T9 Text Input. (n.d.). Retrieved March 30, 2010, from <http://www.t9.com/us/learn/>
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing in-formation. *Psychological Review*, 63(2), 81-97.
- Warm, J. S., Dember, W., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In R. Parasuraman & M. Mouloua (Eds.), *Automation and human performance: Theory and applications*, 183–200. Mahwah, NJ: Erlbaum.