

## ***AN INITIAL ASSESSMENT OF THE SIGNIFICANCE OF TASK PACING ON SELF-REPORT AND PHYSIOLOGICAL MEASURES OF WORKLOAD WHILE DRIVING***

Bruce Mehler<sup>1,2</sup> & Bryan Reimer<sup>1,2</sup>

<sup>1</sup>The Massachusetts Institute of Technology AgeLab

<sup>2</sup>New England University Transportation Center

Cambridge, Massachusetts, USA

E-mail: bmehler@mit.edu

**Summary:** In block A of a simulator study, a sample of 38 drivers showed a stepwise increase in heart rate and skin conductance level (SCL) from single task driving and across 3 levels of an auditory presentation – verbal response dual task (n-back), replicating findings from on-road research. Subjective ratings showed a similar stepwise increase, establishing concurrent validity of the physiological indices as measures of workload. In block B, varying the inter-stimulus interval in the intermediate 1-back level of the task resulted in a pattern across self-report workload ratings, heart rate, and SCL suggesting that task pacing may influence effective workload. Further consideration of the impact of task pacing in auditory-verbal in-vehicle applications is indicated.

### **INTRODUCTION**

There is a growing recognition that the cognitive demands on attention from cell-phone conversations, audio translations of text messages, and voice-command system interactions are situations where a driver's attention can be diverted even while their eyes may remain directed toward the roadway. These scenarios are often described as conditions that potentially result in a loss of situational awareness, inattention blindness, or as situations of "look-but-fail-to-see" (Kass, Kerstan, & Stanny, 2007; Recarte & Nunes, 2003; Strayer, Drews, & Johnston, 2003). Current efforts to understand distraction often focus on specific technologies as opposed to underlying principles. While the multiple resource theory (Wickens, 2008) and the 4-D multiple resource model provide a strong conceptual framework for predicting overload, other models of driver attention may also be applicable. Across these frameworks a number of gaps appear in our understanding of how different dimensions of workload affect driver performance, visual attention, arousal and behavior.

One research gap relates to the pace with which an individual interacts with a vehicle system. This is not to say that the overall topic of task pacing has not been addressed. For instance, the Alliance of Automotive Manufacturers' Driver Focus guidelines (2006) included principles 3.4 and 3.5 related to "task pacing". However, these principles were enumerated largely as general concepts and not developed in any depth. In a 2007 NHTSA report on the characteristics of voice based interfaces (Ranney, Mazzae, Baldwin, & Salaani, 2007), Tom Ranney and colleagues concluded that "Additional research is recommended to explore the effects of higher system error rates and task pacing (self-paced vs. externally-paced) on the interference induced by secondary task performance". At the present time, there is little research available on which to develop guidelines on pacing at a more detailed level. The knowledge gap in this area is evident in the draft NHTSA visual manual distraction guidelines (2012), which did not include any principles related to task pacing.

Extensive research has been undertaken on the use of a delayed digit recall task (n-back) as a method for inducing, and studying the impact of, graded levels of cognitive demand during driving (Mehler, Reimer, Coughlin, & Dusek, 2009; Mehler, Reimer, & Coughlin, 2012; Reimer, 2010; Reimer & Mehler, 2011; Reimer, Mehler, Wang, & Coughlin, 2012). In the format developed by our group, this task provides participants with a fixed auditory demand, consisting of listening to a randomly ordered series of numbers (0-9) presented one at a time, and requires verbalization of the  $n^{\text{th}}$  preceding number to the one just presented. In the case of the 0-back task, this corresponds to simply repeating each number as it is presented. In the 1-back task, this means holding a number in memory, waiting for the next number to be presented and entering it into memory, and verbalizing the previous number while continuing to hold the most recent number in memory. The 2-back extends upon the 1-back by requiring that the participant holds the most recent two numbers in memory. The vocal demands of this task are relatively consistent, i.e. the 1-back requires one less vocalization than the 0-back and the 2-back requires one less vocalization than the 1-back. Consequently, the task largely represents a manipulation of the level of demand on working memory.

As discussed in Mehler et al. (2012), the pacing of this task may have a relationship to effective workload. In previous research, a fixed inter-stimulus interval of 2.25 seconds was employed. The present study aims to evaluate the significance of this aspect of pacing by presenting the 1-back task under a range of fixed pacing intervals and examining the impact on physiological measures of workload. In addition, the three “classic” levels of the n-back (0, 1, and 2) are considered along with a new condition, a “blank-back”, that only requests listening to the auditory prompts and thus has no intrinsic memory requirement. Finally, the research on this task is extended to include self-reported workload assessed at the completion of each demand level.

## **METHOD**

### **Subjects**

Participants were required to be active, experienced drivers, defined as driving on average 3 or more times a week and having held a valid driver’s license for 3+ years. They were drawn from community volunteers in the greater Boston area who responded to online, print advertisements, or referrals. Compensation of \$30 was provided. The study was approved by MIT’s institutional review board. The data reported here were collected from 38 participants (23 males) aged 20-31 years. The mean age was 23.4 years (SD=2.8) for males and 24.7 years (SD=3.1) for females.

### **Apparatus**

*Driving simulator.* The study was conducted in a driving simulator consisting of a fixed base, full cab Volkswagen New Beetle situated in front of an 8’ by 6’ (2.44m by 1.83m) projection screen positioned 76” (1.93m) in front of the mid-point of the windshield. This provided approximately a 40 degree view of the virtual world at a resolution of 1024 x 768 pixels. Graphical updates were generated at a minimum frame rate of 20 Hz by STISIM Drive version 2.08.02 (Systems Technology, Inc., Hawthorne, CA) based upon a driver’s interaction with the steering wheel, brake and accelerator. A MEDAC System/3 physiological monitoring unit

(NeuroDyne Medical, Cambridge MA) was sampled at a rate of 250 Hz. to obtain heart rate (modified lead II EKG configuration) and electrodermal activity (skin conductance level) following placements described in (Mehler et al., 2012). Audio tasks and instructions were pre-recorded and played through the vehicle sound system.

*Driving environment.* The simulation consisted of a divided highway with two lanes in each direction plus a 2 foot (0.61 m) shoulder. Lane width was 15 feet (3.62 m) and posted speed limit was 65 mph (104.6 km/h). Typical traffic events on the virtual highway included passing vehicles, lane changes, and slow downs. The average traffic density in the virtual scenario was set at 23 vehicles/mile (14.3/km). Average traffic speed for vehicles in the left lane was set equal to the posted speed limit of 65 mph (104.6 km/h) and 5 mph slower (96.5 km/h) for the right lane.

*Secondary task (n-back).* A description of the form of the n-back task used in this study is provided in the introduction. The present investigation adds a “blank-back” condition (in addition to the 0, 1, and 2-back) in which the participant is instructed to simply listen to the series of single digit numbers as they are presented; no memorization or verbalization is required. See Mehler, Reimer and Dusek (2011) for additional background on the development, validation, and recommended participant training on the task.

## **Procedure**

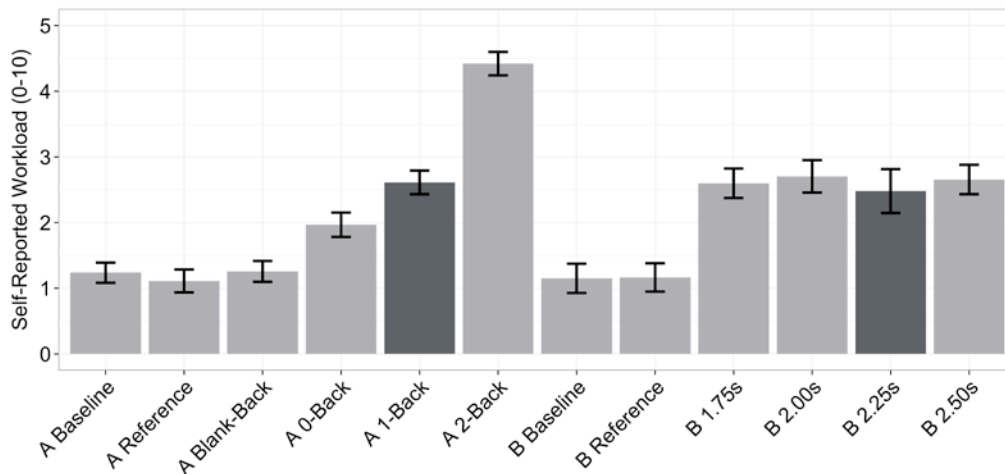
Participants read and signed an informed consent and eligibility was verified by an interview. Participants were trained on the n-back task following established procedures (with the addition of training in the blank-back task) and performance incentives for the simulator detailed in Mehler, Reimer and Dusek (2011). A brief drive of 2.65 miles (approximately 4 minutes) followed to provide a degree of familiarization with the simulator environment. Participants were then instructed to pull over to the side of the roadway and stop the car. Additional practice with each of the 4 levels of the n-back was provided by an automated system.

The experimental portion of the study was divided into two blocks, A & B. Each block was composed of 6 task periods. At the start of each task period, a driver accelerated the simulator at their own pace up to highway speed. When a participant reached 55 mph, a count-down of an additional 60 seconds of adaption time began, after which recorded instructions for the content of the task interval of the period were presented. The first period in each block consisted solely of single task driving (A Baseline or B Baseline) and the assessment interval lasted 60 seconds; all other task intervals were 30 seconds in duration. In Block A, the subsequent 5 periods consisted of one instance each of continued single task driving (A Reference), a blank, 0, 1, and a 2-back. The inter-stimulus interval for the n-backs was the standard 2.25 seconds from the start of one utterance and the start of the following utterance used in previous studies. In Block B, the subsequent 5 periods consisted of continued single task driving (B Reference) or one instance each of a 1-back with an inter-stimulus interval of either 1.75, 2.0, 2.25, or 2.5 seconds. Half the participants were exposed to Block A first and half to Block B. Task order within blocks was pseudo-randomly shuffled to control for order effects. Participants stopped the vehicle by the side of the roadway at the end of each period and completed a workload rating scale for the task portion of the preceding period. In the case of single task driving, participants were asked to rate the workload associated with driving alone.

## RESULTS

### Self-reported workload

Participants were asked to rate the workload of each of the twelve task periods on a scale of 0-10 (Figure 1). Subjective workload differed significantly between tasks ( $X^2_{(11)} = 245.2$ ,  $p < .001$ , Friedman test). Sub-tests show that participants rated the four levels of the n-back (blank, 0, 1, and 2-back) in Block A significantly different ( $X^2_{(3)} = 58.2$ ,  $p < .001$ ), with subjective workload increasing monotonically with the objective load level ( $r = 0.57$ ,  $p < .001$ , Pearson correlation). In Block B, differences in mean workload ratings of the 1-back tasks with different inter-stimulus intervals were not statistically significant from one another ( $X^2_{(3)} = 2.59$ ,  $p = .459$ ).



**Figure 1. Mean self-reported workload on a low to high scale of 0 to 10 (with half point resolution) (note that the “A 1-back” and “B 2.25sec” tasks (dark bars) were objectively equivalent and simply conducted in different blocks; error bars represent  $\pm 1$  between-subject SEM)**

### Skin conductance level

Mean skin conductance level (SCL) varied significantly across task periods ( $F_{(11, 407)} = 7.77$ ,  $p < .001$ , ANOVA with repeated measures) (Figure 2). SCL also varied across the four demand levels of the n-back ( $F_{(3, 111)} = 12.8$ ,  $p < .001$ ) in Block A. SCL did not differ across the four different inter-stimulus timings of the 1-back ( $F_{(3, 111)} = 1.3$ ,  $p = .284$ ). Pairwise comparisons between the blank-back and the non-task periods were significantly different ( $t_{(37)} = 3.34$ ,  $p < .01$ ;  $t_{(37)} = 3.07$ ,  $p < .01$ ; comparing blank-back to baseline and reference periods, respectively).

### Heart rate

Mean heart rate varied significantly across task type ( $F_{(11, 407)} = 20.9$ ,  $p < .001$ , ANOVA with repeated measures) (Figure 3). Heart rate varied significantly across the four levels of the n-back ( $F_{(3, 111)} = 46.1$ ,  $p < .001$ ) in Block A and monotonically with objective demand ( $r = .33$ ,  $p < .001$ ). Mean heart rate did not differ significantly across the different stimulus timings used in Block B ( $F_{(3, 111)} = 0.45$ ,  $p = .717$ ).

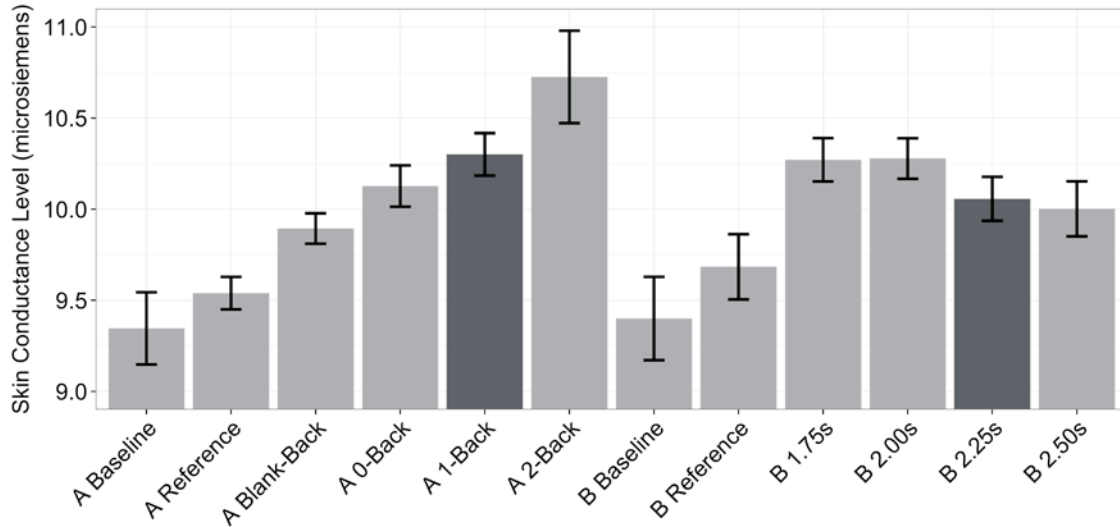


Figure 2. Average skin conductance level across task types (labeling and annotations as in Figure 1)

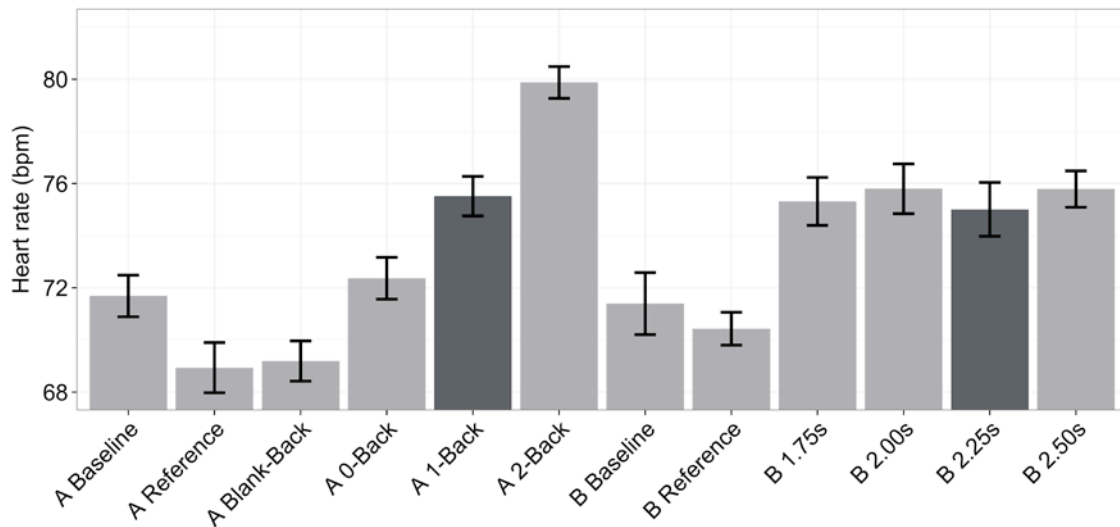


Figure 3. Average heart rate across task types (labeling and annotation as in Figure 1)

## DISCUSSION

As was observed under actual driving conditions (Mehler et al., 2012), heart rate and SCL increased in Block A in a relative stepwise fashion with each increase in the level of objective demand from baseline single-task driving through driving plus engagement with the 0, 1, and 2-back tasks. Concurrent validity of these indices as measures of workload is established in the subjective workload ratings that also show an increase with each demand level. We did not collect subjective data in earlier work since it would require interrupting the flow of experimental conditions to obtain successive self-ratings. The ability to obtain a continuous measure of workload is generally seen as one of the advantages of physiological measures (Brookhuis & de Waard, 2001; Mehler et al., 2009).

Extending previous work on the sensitivity of physiological indices as measures of workload, in the new “blank-back” condition, where participants were asked only to listen to the numbers as they were presented, SCL was found as expected to fall between the initial driving only baseline condition and the low demand 0-back condition. Thus, the blank-back might be classified as a low-low demand condition. Interestingly, it can be observed in Block A that heart rate during the blank-back appears to drop below the level seen during the baseline driving only condition and is essentially indistinguishable from the heart rate observed during the driving only reference period. This may be an example of fractional differentiation, where there is an apparent divergence of heart rate from SCL as an arousal measure under what is known as a “sensory intake” state (Lacey & Lacey, 1974; Mehler et al., 2009). Quietly listening to the series of numbers during the blank-back fits well with the sensory-intake pattern, i.e. heart rate drops or remains low during the intake of auditory information, while SCL increases due to the arousal associated with the request to listen. Conversely, during the 0, 1, and 2-back tasks, both heart rate and SCL increase, in what has been termed a “sensory-rejection” condition, as the driver attempts to ignore extraneous sensory information while holding numbers in working memory. As noted elsewhere (Mehler et al., 2012), this type of patterning demonstrates that while increases in heart rate can fairly reasonably be interpreted as evidence for an increase in arousal, the lack of an increase or a drop in heart rate cannot automatically be taken as evidence for its absence. This is one of the reasons why we also collect SCL data in research applications.

The 2.25 second inter-stimulus interval used in earlier studies was selected based upon limited pilot-testing. The pattern of heart rate and SCL data in Block B suggests that this interval may in fact be a fairly optimal selection. Participant observations gave the impression that shorter 1.75 and 2.0 second intervals increased the pressure of the task and this is coincident with nominally higher heart rate and SCL. Drawing the pacing out to 2.5 seconds was also associated with nominally higher reported workload and heart rate, although not SCL. In Reimer et al. (2012), it was found that in the high demand 2-back condition drivers tended to respond faster. As the amount of information to be maintained in working memory increases, there may be pressure to “dump” load from working memory as soon as possible. Thus, being pressured to respond quickly and being forced to hold a response longer than is optimal may both be potential sources of added workload. Similar factors may be present when task pacing is controlled by a DVI such as in a multi-step voice-command interaction. Pacing may be considered at the broad level of user-paced vs. fixed paced, or at the more micro-level of system-paced intervals as considered here. Both areas appear worthy of further study. It is important to consider that an optimal pacing interval is likely to vary by individual and thus a study of self-paced interactions should be considered in addition to further basic research on fixed intervals considered here.

## ACKNOWLEDGMENTS

We gratefully acknowledge the supported of The Santos Family Foundation and the US Department of Transportation’s Region I New England University Transportation Center.

## REFERENCES

Alliance of Automobile Manufacturers (2006). *Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems*. Driver Focus Telematics Working Group, Washington, D.C.

- Brookhuis, K.A. & de Waard, D. (2001). *Assessment of drivers' workload: Performance and subjective and physiological indexes*. In P.A. Hancock and P.A. Desmond (Eds.), *Stress, Workload, and Fatigue*, Mahwah, NJ: Lawrence Erlbaum Associates, pp. 321-333.
- Kass, S.J., Kerstan, K.S., & Stanny, C.J. (2007). Effects of distraction and experience on situation awareness and simulated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(4), 321-329.
- Lacey, B. C., & Lacey, J. I. (1974). Studies of heart rate and other bodily processes in sensorymotor behavior. In P. A. Obrist, A. H. Black, J. Brener & L. V. DiCara (Eds.), *Cardiovascular psychophysiology*, Chicago, IL: Aldine Publishing Company, pp. 538-592.
- Mehler, B., Reimer, B. & Coughlin, J.F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Human Factors*, 54(3), 396-412.
- Mehler, B., Reimer, B., Coughlin, J.F., & Dusek, J.A. (2009). The impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record*, 2138, 6-12.
- Mehler, B., Reimer, B. & Dusek, J.A. (2011). *MIT AgeLab delayed digit recall task (n-back)*. MIT AgeLab White Paper Number 2011-3B. MIT, Cambridge, MA.
- Mehler, B., Reimer, B., Pohlmeier, A.E., & Coughlin, J.F. (2008). The association between heart rate reactivity and driving performance under dual task demand in late middle age drivers. *Advances in Transportation Studies an International Journal. Special Issue*, 53-70.
- NHTSA (2012). *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*. National Highway Traffic Safety Administration, U.S. DOT, Washington, DC.
- Ranney, T.A., Mazzae, E.N., Baldwin, G.H.S., & Salaani, M.K. (2007). *Characteristics of voice-based interfaces for in-vehicle systems and their effects on driving performance*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.
- Recarte, M.A. & Nunes, L.M. (2003). Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of Experimental Psychology: Applied*, 9(2), 119-137.
- Reimer, B. (2009). Cognitive task complexity and the impact on drivers' visual tunneling. *Transportation Research Record*, 2138, 13-19.
- Reimer, B. & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics*, 54(10), 932-942.
- Reimer, B., Mehler, B., Wang, Y., & Coughlin, J.F. (2012). A field study on the impact of variations in short term memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors*, 54(3), 454-468.
- 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.D. (2008). Multiple resources and mental workload. *Human Factors*, 50(3), 449-455.