

***HOW DEMANDING IS “JUST DRIVING”?  
A COGNITIVE WORKLOAD – PSYCHOPHYSIOLOGICAL REFERENCE EVALUATION***

Bruce Mehler, Bryan Reimer

MIT AgeLab & N.E. University Transportation Center, Cambridge, MA USA

Email: bmehler@mit.edu

**Summary:** Physiological arousal, measured as heart rate and skin conductance level, was recorded during single-task highway driving (just driving), while driving and interacting with several voice-based and visual-manual infotainment user interfaces, while driving and engaging in multiple levels of a cognitive workload reference task (n-back), and while engaging in the same cognitive workload reference task under single-task (non-driving) conditions. Single-task highway driving was found to produce a level of physiological arousal in the same range as that of the relatively highly demanding 2-back task under non-driving conditions. While continuing innovations such as automatic transmission, power steering, as well as climate control, sound proofing and other comfort features, have reduced the overt demands of driving, these findings suggest that the remaining demand on resources during what has been thought of as “just driving” may be higher than many realize. The extent to which various implementations of longitudinal and lateral control driver assistance features being introduced change this dynamic is largely an open question.

## **INTRODUCTION**

For experienced drivers, highway driving under nominal conditions (low to moderate traffic, good weather, etc.) may be experienced as an over-learned skill that seemingly requires little effort. However, is that actually the case? How demanding on the body’s resources is the act of “just driving”?

This paper considers this question by leveraging data that was collected as part of a series of on-road studies examining the relative demand placed on drivers by a range of voice-based and visual-manual infotainment tasks in production vehicles (e.g., Reimer, Mehler, Dobres & Coughlin, 2013; Mehler et al., 2014; Mehler, Reimer, Dobres, & Coughlin, 2015; Mehler, Reimer, Dobres, Foley & Ebe, 2016). In addition to other dependent variables, physiological arousal levels were monitored as objective indicators of workload. Physiological indices such as heart rate are known to be responsive to both physical and cognitive demands (e.g. Szabo, Péronnet, Gauvin, & Furedy, 1994) and can be leveraged to provide a form of net load assessment. Participants also engaged in multiple levels of a working memory task (detailed below) to provide a scaled reference for their physiological arousal level during various periods of the drive. In the dataset reported here, the cognitive reference task was also administered under static (non-driving) conditions in a parking lot at the end of the drive. This provided the opportunity to compare these scaled levels of physiological arousal under single task conditions to the levels observed when adding such cognitive load during the driving task, as well as to consider how the demand of “just driving” compares to these reference values.

## **METHOD**

### **Participants**

Recruitment targeted a community sample from the greater Boston area with a younger (20-29 yrs.) and an older (60-69 yrs.) age cohort. Recruitment methods and the study protocol were approved by MIT's institutional review board. Consistent with other studies in this project stream, participation required that individuals meet the following criteria: a) have held a valid driver's license for more than three years, b) drive on average three or more times per week, c) be in self-reported reasonably good health for their age and meet a set of health exclusion criteria, d) clearly understand and speak English, e) no involvement as a driver in a police reported accident in the past year, f) and not actively using any medications causing drowsiness (see Mehler, Reimer, Dobres, & Coughlin (2015) Appendix F for full details). Participants who completed the intake process, initial secondary task training, and went on-road were paid \$150.

### **Apparatus & Measures**

The study vehicle was a 2015 Toyota Corolla with Entune Premium Audio with navigation. It was instrumented with a custom data acquisition system for time synchronized recording of data including: a) vehicle information via the controller area network (CAN) bus, b) a Garmin 18X Global Positioning system (GPS) unit, c) a MEDAC System/3™ physiological monitoring unit to provide an electrocardiogram (EKG) signal for calculating instantaneous heart rate values and skin conductance level (SCL), d) multiple video cameras to capture the driver's face, their interactions with the vehicle's steering wheel and center console, the forward roadway, and a rear roadway view, e) a wide area microphone to capture driver speech and audio from the vehicle's speech system. Physiological signals were recorded at 250 Hz to support EKG feature extraction for heart beat interval detection. Subjective workload was assessed via questionnaire using a single global rating per task type on a 0 (low) to 10 (high) scale that allowed for half-interval ratings (21 points). The instruction set and scale have been demonstrated to produce ratings consistent with relative rankings of global scores obtained using the NASA Task Load Index (see Mehler et al., 2016).

### **Secondary Tasks**

The full experimental protocol considered a range of voice-command based tasks using the in-vehicle user interface along with established visual-manual reference tasks (radio tuning) and auditory-vocal-cognitive calibration reference tasks (multi-level n-backs). The tasks and protocol for task presentation were identical to those detailed in Mehler et al. (2015) with the exception that an additional set of n-back tasks was presented in a parking lot at the end of the study under static (non-driving) conditions. In brief, the voice-based tasks consisted of full address destination entry into the navigation system, point of interest (POI) destination entry, canceling navigation, and phone contact calling. The visual-manual radio tasks consisted of single button press preset station selection (Radio Easy) and multi-button press and manual rotation of the tuning knob to select specified stations (Radio Hard) in conformance with the NHTSA (2013) radio tuning reference task. The n-back was a delayed digit recall, working memory task in an audio presentation / verbal response format presented at 0, 1, and 2-back levels. See Mehler,

Reimer, and Dusek (2011) for background, training materials, and links to audio files for research use. The numbers 0–9 were presented one at a time, once each, in a random sequence over a 30 second interval. At the lowest demand level (0-back), participants were to respond by simply immediately repeating out loud each number as it was presented. At the medium level (1-back), participants had to enter each new item in memory and respond verbally with one number back in the sequence. At the most difficult level (2-back), participants had to hold the two previous items in memory in the correct order, take in the new item, and respond with the number two items back in the sequence. The presentation order for three difficulty levels was randomly distributed across the sample.

## **Procedure**

As previously noted, the protocol for training and task presentation was identical to that detailed in Mehler et al. (2015) with the exception of the final set of n-back tasks that was presented under non-driving conditions in a parking lot at the end of the study. Participants read and signed an informed consent, received initial in-lab training, and physiological sensors were attached. A modified lead II configuration was used for EKG recording; the negative lead was placed just under the right clavicle (collar bone), ground under the left clavicle, and positive lead on the left side over the lower rib. The skin was cleaned with isopropyl alcohol and standard pre-gelled silver/silver chloride disposable electrodes (Vermed A10005, 7% chloride wet gel) applied. Skin conductance was measured utilizing a constant current configuration. Non-polarizing, low impedance, thin gold plated electrodes allowed electrodermal recording without the use of conductive gel. Sensors were taped to the underside of the outer segments of the middle fingers of the left hand so that the right hand could be freely used to interact as needed with the vehicle infotainment system. Approximately 20 minutes of training on in-vehicle secondary tasks, including an n-back practice set, took place in a parking lot. This was followed by approximately 30 minutes of adaptation to active driving of the vehicle prior formal task assessment. The sections of highway for the periods of driving considered here consisted of three travel lanes in each direction, were bordered largely by forest, and had a posted speed limit of 65 mph.

## **Data Reduction & Analysis**

Single task driving reference periods were calculated for 2 minutes of “just driving” prior to a recorded audio message indicating the start of a new task period. The four periods used for this purpose (for 8 minutes total) were collected just prior to prompts indicating the start of the n-back, destination address entry, contact phone calling, and POI entry tasks. Metrics were calculated and mean values across the reference periods were used as an overall just driving reference. For secondary task periods, values for each dependent measure were calculated per trial and mean values across trials were used for analytic purposes.

R-wave peaks in the EKG signal were identified to determine inter-beat intervals and for calculating instantaneous heart rate using software developed at the MIT AgeLab. Consistent with existing standards (Task Force, 1996), automated detection of R-wave peaks were visually reviewed and misidentified and irregular intervals manually corrected. Another MIT AgeLab developed data processing package removed high-frequency noise in the skin conductance signal as per Reimer and Mehler (2011) and any substantive identified motion artifacts manually edited.

Statistical analyses were performed in R (R Core Team, 2014). Due to the non-normal distribution of the data and/or the use of ratio data (percentages) for several dependent measures, in many cases non-parametric statistics - the Wilcoxon signed rank test and the Friedman test - were used (similar to the t-test and repeated-measures ANOVA, respectively). These tests have been shown to be more robust against Type I error in cases where data are non-normal (Conover & Iman, 1981; Friedman, 1937).

## RESULTS

To be included in the analysis sample, participants had to demonstrate the ability to complete each task under controlled conditions in the parking lot and engage in each task during the drive. Cases were excluded if non-optimal weather conditions (e.g. heavy rain) or dense traffic was encountered. The analysis sample consisted of 12 participants, balanced by gender and equally distributed across the younger (20-29 yrs.) and older (60-69 yrs.) age groups (see Table 1).

Table 1. Summary age statistics by gender

Gender	Mean Age	SD	Minimum	Maximum	N
Female	44.0	23.9	21	69	6
Male	43.5	25.4	20	68	6
					12

### Self-Reported Workload

Participants rated how much workload they experienced while engaged in each task while driving. Summary statistics are presented graphically in Figure 1. Ratings were not collected for the supplemental n-back task periods while vehicle was parked at the conclusion of the drive. Self-reported workload ratings differed significantly across the tasks ( $X^2(9) = 71.3, p < 0.001$ ).

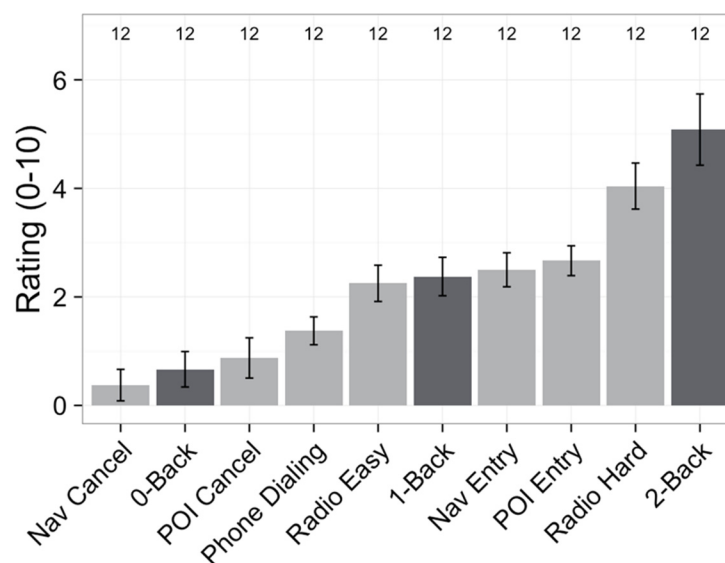
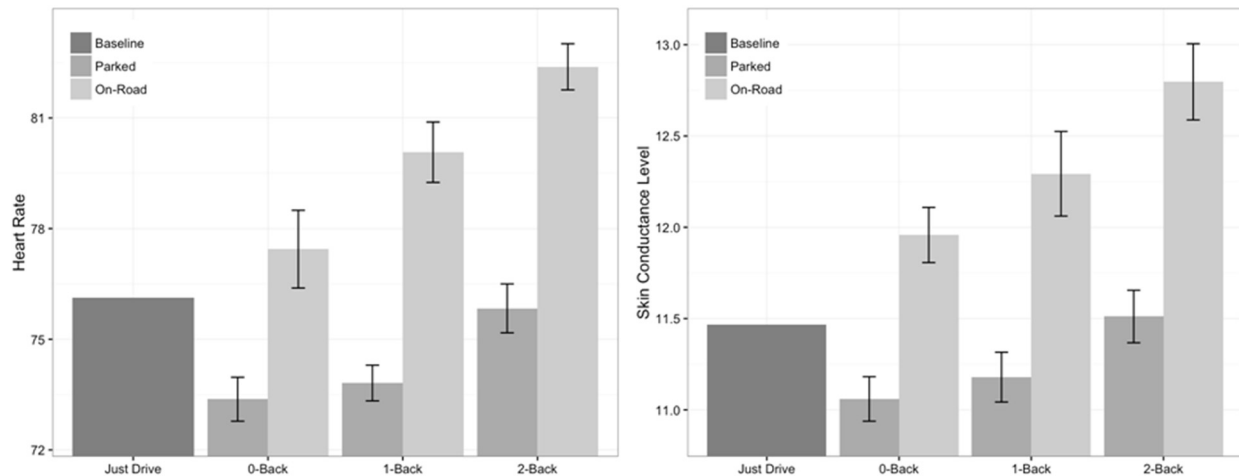


Figure 1. Self-reported workload ratings in ascending order from low (0) to high (10) for each secondary task under study while driving. Bars represent mean values while error bars represent the mean-adjusted standard error. Darker bars indicate the n-back cognitive loading tasks undertaken while driving.

## Physiological Data

As expected, peripheral physiological arousal measures (heart rate and SCL) were higher during the n-back tasks when participants were driving on the highway vs. under static conditions when the vehicle was parked (see Figure 2). It can also be observed that within both the static and driving conditions, heart rate and SCL values increased with increasing difficulty level of the n-back. Considering the on-road data, there was a significant main effect of heart rate across baseline driving and the three added n-back task demand levels ( $X^2(3) = 16.7, p < 0.001$ ). There was also a significant main effect for SCL ( $X^2(3) = 8.7, p < 0.034$ ).



**Figure 2. Mean heart rate in beats per minute (left) and skin conductance level (SCL) in microsiemens (right) during the “just drive” reference period (wide dark bar) and during three levels of the n-back auditory-vocal cognitive workload reference task while driving (light bars) and while parked (medium dark bars). Error bars represent the mean-adjusted standard error.**

## DISCUSSION

The overt demand characteristics of the three levels of the n-back working memory task suggest that workload should increase over just driving with increasing levels of “n” as drivers engage with the tasks. In line with previous research with these tasks (Mehler, Reimer, & Coughlin, 2012; Reimer, Mehler, Dobres, & Coughlin, 2013), self-reported workload ratings and peripheral physiological measures (heart rate and SCL) collected during the four conditions (just driving, and driving plus 0, 1, and 2-back) indicate that effective cognitive workload / arousal did in fact increase in a step-wise fashion with each demand level. It can be observed that this is a rather robust effect in that it appears in this relatively modest sized sample (N=12) under the variable conditions of actual highway driving.

The scaling of self-reported workload for the n-back reference tasks also align well with previous work (Reimer et al. 2013; Mehler et al. 2015, 2015). The low demand 0-back is ranked in the same low workload range as the two voice-based route cancelation tasks that are relatively easy to execute with the implementation in this vehicle. At the other end of the rankings is the high demand 2-back task with the visual-manual radio tuning reference task falling just below this surrogate task reference point.

Interestingly, the nominal magnitude of the difference between the low-demand 0-back and medium demand 1-back are not as distinct when presented as single task challenges in the static parking lot condition compared to the relative impact on physiological arousal when presented as secondary tasks during active driving. Thus, the same incremental increase in objective demand may have greater relative impact on an operator under multi-tasking conditions than is observed under single task conditions. This is worth keeping in mind when the demands of a vehicle HMI are being assessed under test-bench conditions.

The motivation for adding the set of multi-level n-back tasks under the static parked condition was to provide a means for assessing the extent to which potential movement artifacts did or did not impact any underlying signal in a new physiological recording device that was being tested under the driving conditions reported here. A side result of this assessment was the opportunity to observe how the physiological arousal indicators of workload during highway driving compared to that of the n-back reference task under single-task, non-driving conditions. As might be expected, the physiological arousal level associated with the very low demand 0-back task was lower than that seen during driving. However, single-task highway driving was found to produce a level of physiological arousal in the same range as that of the relatively highly cognitively demanding 2-back task. Thus, while highway driving, as an over-learned task for experienced drivers, may generally seem to be relatively low-demanding, the level of actual workload associated with the task may be significantly greater than many realize.

It seems likely that a significant portion of the workload of just driving (objectively characterized in physiological arousal) is associated with the level of monitoring and vigilance required to safely operate the vehicle. With this in mind, it seems worth considering that while SAE (2018) level 1 and 2 driver assistance systems reduce some of the manual input required to operate a vehicle, the monitoring demands for safe supervision may, in some ways, increase. Consequently, the extent to which various implementations of longitudinal and lateral control driver assistance features result in an effective net reduction of overall workload in this dynamic is something of an open question.

## **ACKNOWLEDGMENTS**

Acknowledgement is extended to the Toyota Collaborative Safety Research Center (CSRC) which provided the primary funding for the larger project from which this data is drawn. Supplemental / matching support was provided by the U.S. Department of Transportation's Region I New England University Transportation Center at MIT. This work would not have been possible without the support of AgeLab staff including: Hillary Abraham, Dan Brown, Jonathan Dobres, Alea Mehler, Hale McAnulty, Tom McWilliams, and Anthony Pettinato in study management, participant recruitment, and data collection, reduction and coding support.

## **REFERENCES**

- Conover, W. J., & Iman, R. L. (1981). Rank transformations as a bridge between parametric and nonparametric statistics. *American Statistician*, 35(3), 124–129.
- Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, 32(200), 675–701.

- Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J. & McCartt, A. (2016). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*, 59(3), 344-367.  
doi:[10.1080/00140139.2015.1081412](https://doi.org/10.1080/00140139.2015.1081412)
- 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, 396–412.
- Mehler, B., Reimer, B., Dobres, J., Foley, J., & Ebe, K. (2016). Additional findings on the multi-modal demands of production level “voice-command” interfaces. *SAE Technical Paper 2016-01-1428* and presentation at *SAE World Congress*, Detroit, MI.
- Mehler, B., Reimer, B., Dobres, J., McAnulty, H., Mehler, A., Munger, D., & Coughlin, J.F. (2014). *Further evaluation of the effects of a production level “voice-command” interface on driver behavior: replication and a consideration of the significance of training method*. MIT AgeLab Technical Report No. 2014-2. Massachusetts Institute of Technology, Cambridge, MA. Hyperlink: [265594707](https://doi.org/10.265594707)
- Mehler, B., Reimer, B. & Dusek, J.A. (2011). *MIT AgeLab delayed digit recall task (n-back)*. MIT AgeLab White Paper Number 2011–3B. Massachusetts Institute of Technology, Cambridge, MA.
- Mehler, B., Reimer, B., Dobres, J., & Coughlin, J.F. (2015). *Assessing the Demands of Voice Based In-Vehicle Interfaces: Phase II Experiment 3 - 2015 Toyota Corolla (2015b)*. MIT AgeLab Technical Report No. 2015-14. Massachusetts Institute of Technology, Cambridge, MA. doi: 10.13140/RG.2.1.2244.3921
- National Highway Traffic Safety Administration. (2013). *(Issued Guidelines) Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (Docket No. NHTSA-2010-0053)*. Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.R-project.org/>
- 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., Dobres, J. & Coughlin, J.F. (2013). *The effects of a production level “voice-command” interface on driver behavior: reported workload, physiology, visual attention, and driving performance*. MIT AgeLab Technical Report No. 2013-17A (November 18, 2012). Massachusetts Institute of Technology, Cambridge, MA.
- SAE (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles, *SAE Standard J3016*, USA.
- Szabo, A., Péronnet, F., Gauvin, L., & Furedy, J. J. (1994). Mental challenge elicits “additional” increases in heart rate during low and moderate intensity cycling. *International Journal of Psychophysiology*, 17(3), 197-204.
- Task Force (1996). Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, 17, 354-381.