High Cognitive Load Assessment in Drivers through Wireless Electroencephalography and the Validation of a Modified N-Back Task

Dengbo He, Birsen Donmez, *Senior Member*, *IEEE*, Cheng Chen Liu, and Konstantinos Plataniotis, *Fellow*, *IEEE*

***Abstract*— This paper explores the influence of high cognitive load on driver's Electroencephalography (EEG) signals collected from two channels (Fp1, Fp2) through a wireless consumer-grade system. Although EEG has been used in driving research to assess cognitive load, only a few studies focused on high load and they used research-grade systems. Recent advancements allow for less intrusive and more affordable systems. As an exploration, we tested the feasibility of one such system to differentiate three levels of cognitive taskload in a simulator study. Thirty-seven participants completed a baseline drive with no secondary task and two drives with a modified version of the n-back task (1-back, 2-back). The modification removed the verbal response required during task presentation to prevent EEG signal degradation, with the 2-back expected to impose higher load than 1-back. Another objective of this study was to validate that this modified task increased cognitive load in the expected manner. The modified task led to significant trends from baseline to 1-back, and from 1-back to 2-back in heart rate, galvanic skin response, respiration, variability in horizontal gaze position, and pupil diameter, all in line with previous driving studies on cognitive load. Further, the EEG system was sensitive to the modified task, with the power of alpha band decreasing significantly with increasing n-back levels (baseline vs. 1-back: 0.092 Bels on Fp1, 0.179 on Fp2; 1-back vs. 2-back: 0.209 on Fp1, 0.147 on Fp2). Thus, a consumer-grade EEG system has the potential to capture high levels of cognitive load experienced by drivers.**

***Index Terms*—Driver assistance systems, Driving performance, N-back task, Physiological measures, Electroencephalography**

I. Introduction

DRIVING can be mentally demanding, especially under certain circumstances such as bad weather and complex traffic conditions. Activities secondary to driving, e.g., the use of in-vehicle infotainment systems and smart phones, can also claim cognitive resources. Although visual-manual secondary tasks are especially detrimental to safety [1], tasks that are auditory-verbal are also of concern as they are becoming more common with the rise of voice-command interfaces within the vehicle. Therefore, there is a need to also study the effects of auditory-verbal secondary tasks on drivers' cognitive load, a multidimensional construct representing the load that performing a particular task imposes on the drivers' cognitive system [2]. What the driver experiences, i.e., cognitive load, depends on taskload (based solely on task characteristics) as well as the individual driver characteristics and the cognitive capacity that the driver allocates to different tasks [2]. Several simulator and on-road studies indicate that auditory-verbal secondary tasks impair drivers' visual scanning behaviors and driving performance [3, 4]. Although drivers can moderate their cognitive load to some extent, such as by reducing their speed, avoiding lane changes, and increasing their headway [5, 6], these actions may not be sufficient to fully compensate for the external demands experienced by the drivers. In-vehicle information systems and advanced driver assistance systems can help drivers to better modulate their cognitive load through real-time assessment of cognitive load, e.g., [7], and through ensuing interventions, e.g., locking drivers out from cell phone use when they are detected to be overloaded.

Various measures have been used to estimate cognitive load experienced by drivers. These measures can be categorized into four groups: a) physiological, such as Electroencephalogram (EEG), Electrocardiography (ECG), galvanic skin response (GSR), and respiration; b) eye tracking, such as blink rate and gaze position; c) performance-based, such as vehicle speed; and d) subjective, such as NASA Task Load Index (NASA-TLX). Table 1 provides a summary of example cognitive load measures and their response to increased external cognitive taskload, with results from driving studies cited when available.

It is widely agreed that no single measure alone can provide sufficient information to estimate cognitive load [8, 9], and each measure has its pros and cons. For example, subjective measures indicate how the driver feels but do not provide a continuous assessment of load [10]. Driving performance measures are much less intrusive but also less sensitive to low load levels [10]. Eye-tracking captures cognitive load through a reduction in standard deviation (SD) of gaze position [11], but its accuracy is easily influenced by ambient light, which is difficult to control on the road. Physiological measures can provide a continuous assessment of cognitive load, but some

Manuscript received Dec 7, 2017. This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Dengbo He and Birsen Donmez are with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, ON M5S 3G8 Canada. Cheng Chen Liu and Konstantinos Plataniotis are with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 3G4, Canada (e-mail: dengbo.he@mail.utoronto.ca; donmez@mie.utoronto.ca; cheng.liu@mail.utoronto.ca; kostas@comm.utoronto.ca).

require long sampling windows. For example, a reliable estimation of Heart Rate (HR) variability requires at least 2-5 minutes of data collection [12], and the respiration rate of a healthy individual is only around 12 times per minute [13].

Among physiological measures, EEG seems to be a good candidate for detecting drivers' cognitive load given that it requires a small sampling window, e.g., 5 seconds in [8]. EEG records the electrical activity in the brain through electrodes placed on the surface of the scalp [14] and is more appropriate for driver load assessment compared to other neuroimaging methods. For example, Functional Magnetic Resonance Imaging (fMRI) machines are too big and both fMRI and Functional Near-Infrared Spectroscopy (fNIRS) responses are sluggish in time [15]. However, research-grade EEG systems, e.g., with 32 electrodes used in [16, 17], are still too intrusive for in-vehicle applications.

TABLE I
EXAMPLE COGNITIVE LOAD MEASUREMENTS

Measure	Trend with Increased Cognitive Taskload
Physiological	
EEG	Power of alpha band ↓ [8, 14]
	P300 latency ↑ [17]
ECG	HR ↑ [9, 11, 18]
	HR variability ↓ [18]
GSR	↑ [9, 11]
Respiration	Rate ↑ [9]
Eye Tracking	
Gaze position	Periphery/mirror/instrument check rate ↓ [4]
	SD of horizontal position ↓ [11, 19]
	SD of vertical position ↓ [19]
Blink	Rate ↑ [19]
Pupil diameter	↑ [3, 20]
Performance-based	
Vehicle speed	Average ↑ [9] ↓ [11]
	SD ↑ [9] ↓ [11]
Steering wheel	Reversal rate ↑ [11]
Subjective	
NASA-TLX	↑ [4]

Recent advancements in technology has allowed for the development of less intrusive and much more affordable EEG systems, e.g., 4 channels collected wirelessly through a thin head band. The new technology is clearly much less intrusive and more affordable than research-grade EEG systems and is a step toward achieving a monitoring system that may be accepted by drivers eventually. Further, these newer systems may also be adopted by the research community. Thus, as these systems are being further developed, it is important to test their efficacy. For example, it is unclear whether these newer EEG systems can be used to detect different levels of cognitive load experienced by drivers. As an initial step to address this question, we conducted a driving simulator study to evaluate whether a consumer-grade wireless EEG system can differentiate different levels of cognitive taskload experienced by drivers. Through a secondary task, we imposed three levels of cognitive taskload on our participants and investigated whether EEG signals collected from two channels (Fp1: frontal polar left, Fp2: frontal polar right) provided significant differences among these three taskload levels.

Although EEG technology advances fast, a challenge that remains with these systems is the sensitivity of EEG to artifacts [21], for example, facial muscle movements. Thus, in our simulator study, we had to ensure that facial muscle movements were minimized as we collected EEG data from our participants during periods of interest (i.e., different levels of cognitive taskload). A variety of secondary tasks have been employed in previous driving research to increase cognitive load in a controlled manner, such as the n-back task [9, 11], mental arithmetic [4, 8], and an auditory-spatial task [19]. Amongst these tasks, the n-back task is one of the most widely used and established one in the working-memory literature [22]. During the n-back task, a series of items (e.g., letters or numbers) are presented (visually or aurally) to the participants and participants are required to remember and repeat (verbally or manually) the items n-position before the current one [22]. In driving, this task has been implemented as auditory stimulus and continual verbal response [9, 11], and its relation to common in-vehicle tasks have been investigated extensively [23]. In our experiment, we modified this auditory-verbal n-back task to remove the continual verbal response required during auditory stimulus presentation (an artifact of the n-back task paradigm) which would lead to facial muscle movement interference with EEG signals. The effectiveness of this newly proposed n-back modification was validated through data collected on ECG, GSR, respiration, and eye tracking; we compared our findings to the findings of previous driving studies that investigated cognitive load. The development and validation of this modified n-back task is another contribution of this study. Other researchers who may need a similar modification can utilize our version given that we provided a validation for our version.

II. RELATED RESEARCH

A. *Effects of Cognitive Taskload, in particular N-back Task, on Physiology, Eye Tracking, and Vehicle Control*

A variety of tasks and measures have been implemented in previous driving research to assess how drivers' state and performance are affected in the presence of external cognitive tasks. Some of the relevant studies along with their results are presented in Table 1 and they are further detailed below. The relationships between increased cognitive taskload and the various measures reported in these earlier studies were used to validate our n-back task modification.

Liang and Lee [19], conducted a driving simulator study, in which the participants were presented with an auditory-spatial task that simulated high cognitive taskload (e.g., what drivers may experience while interacting with a navigation system); blink rate was found to increase, while SD of horizontal and vertical gaze position decreased. Harbluk et al. [4] asked their participants to perform mental arithmetic operations in an on-road instrumented vehicle study, and found that their participants checked their periphery, mirrors, and instruments less frequently with this added cognitive taskload. In another on-road study, a paced auditory serial addition task was performed by the participants, during which HR was found to increase and HR variability was found to decrease [18]. Other on-road studies found pupil diameter to increase when drivers were asked to perform cognitive tasks secondary to driving [3, 20].

As mentioned earlier, the n-back task has been widely used in the working-memory literature [22]. N-back task is also well established in driving research and has been used in the form of an auditory delayed digit recall task, with auditory stimulus and verbal response [9, 11, 23]. In the simulator [9] and on the road [11], Mehler et al. tested three n-back levels of increasing difficulty: 0-back, 1-back, and 2-back. In both studies, the n-task performance decreased as ‘n’ increased, confirming the increasing difficulty associated with the three task levels. In relation to actual in-vehicle tasks, 1-back received lower subjective workload ratings from participants than an easy radio task (i.e., single button preset selection), whereas 2-back was perceived to be harder than most common in-vehicle tasks (e.g., a navigation entry task fell midway between the ratings received for 1-back and 2-back tasks) [23].

In [9], Mehler et al. found HR to increase with increasing levels of n-back difficulty. GSR and respiration rate increased from baseline to 0-back and from 0-back to 1-back, with no significant change from 1-back to 2-back, suggesting a plateau for GSR and respiration rate measures at higher levels of n-back taskload. Both speed and SD of speed increased from 1-back to 2-back. In [11], as taskload increased, both HR and GSR increased. Speed and SD of speed were higher in the baseline level compared to all task levels, with speed being further reduced with the 2-back task. Increasing levels of taskload decreased the SD of horizontal gaze position, except there was no difference between 1-back and 2-back levels.

B. *EEG for Assessing High Cognitive Load*

EEG has been used widely in non-driving domains to assess working-memory load. For example, [24] conducted basic research by utilizing EEG with the n-back task performed on a computer. EEG was also utilized to assess high cognitive load in more practical settings, such as aviation [14]. Suppression of the power of alpha band in frontal [25, 26] and parietal [27, 28] areas have been observed with increasing n-back levels in visual-manual n-back task performance in single task situations (i.e., when the n-back task is the only task that the participant conducts). An increase in the power of the theta band has also been observed in the frontal area in similar study settings [26, 27, 29].

Relevant research on multi-tasking situations, in particular research in the driving domain, appears to be more limited, and to focus mostly on primary and secondary tasks that are both visual-manual in nature, e.g., [30], [8]. Lei and Roetting [31] found a decrease in the power of the alpha band in the parietal area and an increase in the power of the theta band in the frontal area in a driving simulator study, when the participants conducted a visual-manual n-back task. Other driving studies we identified that utilized visual-manual secondary tasks (i.e., [16], [32]) did not investigate how specific EEG responses were affected with added taskload, but rather used all EEG signals collected in machine learning algorithms to classify driver state.

Strayer et al. [17] is the only driving study we could identify that utilized EEG with auditory-verbal secondary tasks. Through event-related potential (ERP), P300 peak latency was identified to be sensitive to external cognitive taskload in the laboratory in front of a computer; however, this measure became unreliable in the driving simulator and on the road in an instrumented vehicle. P300 amplitude on the other hand was not sensitive to added cognitive taskload in the computer setting, but showed some sensitivity to added cognitive taskload in the simulator and on the road. Although ERP showed reactivity to external taskload, it is not suitable for cognitive load detection in real time in uncontrolled settings, given that it relies on a response to a specific stimulus (e.g., detection response task) that needs to be identified prior to measurement [33].

Overall, the review of the literature on EEG and driving revealed a significant research gap in the use of EEG to detect cognitive load experienced by drivers. Further, the driving studies reported in this section, which utilized EEG mainly for visual-manual secondary tasks, all had research-grade systems. As mentioned earlier, there are less intrusive and more affordable consumer-grade systems. These systems have the potential to be implemented in vehicles in the future with further development and they may also be adopted by the research community. We could identify only one driving study that used a consumer-grade system, but this study focused on vigilance [34]. Our study detailed below is the first to investigate the reactivity of a consumer-grade EEG system (data from Fp1 and Fp2 channels; both in the frontal area) to external cognitive taskload placed on drivers. In particular, we examined the power of alpha and theta bands. Given the findings of the studies cited above, we expected the power of the alpha band in the frontal area to decrease and the power of the theta band in the frontal area to increase with increasing cognitive load experienced by drivers. However, it should be noted that although these studies were the closest to our task paradigm that we could identify in the existing literature, their task paradigms were still different than ours (i.e., they focused on visual-manual n-back tasks). Therefore, it was also possible that our results could be different than their results.

As also mentioned earlier, we had to modify the n-back task commonly used to study cognitive load in driving, in order to ensure that the EEG signal quality was not degraded by verbal responses that are an artifact of this commonly used n-back task paradigm. Given that the task was altered, another objective of our simulator study was to validate that this modification worked as expected.

III. Driving Simulator Experiment

A driving simulator experiment was conducted with three cognitive taskload conditions in a within-subject design:

baseline (no external secondary task), lower cognitive taskload (1-back task), and higher cognitive taskload (2-back task). Each condition was completed in a separate drive with the order of the three drives counterbalanced across participants.

A. Participants

Participants were recruited through campus and online posts and were required to drive at least several times per month. To improve eye tracking quality, participants were also required to be able to drive without glasses (contact lenses were allowed). Further, they were screened for proneness to simulator sickness. Thirty-seven drivers (18 males and 19 females) with an average age of 26.4 (SD: 4.3, Min: 20, Max: 35) completed the study. Our sample size was larger than that of most relevant research that found effects of external cognitive taskload on driver performance, eye tracking, and physiology [e.g., 3, 4, 8, 18, 19, 20]. The average number of years since our participants obtained their first driving license was 8.5 (SD: 3.5, Min: 2.6, Max: 15.6). Compensation was C$12 per hour, and participants were told that they could receive a bonus of up to C$14 based on their secondary task performance as an incentive for engaging in the n-back task. The experiment took approximately 2.5 hours and all participants were paid the full bonus amount regardless of their performance.

B. Apparatus

The driving simulator used is a NADS miniSim^{TM} (Fig. 1), a fixed-based simulator with three 42-inch screens, creating a 130º horizontal and 24º vertical field of view at approximately 1.2 m viewing distance. The centre screen displays the left and centre parts of the windshield; the right screen displays the rest of the windshield, the rear-view mirror, and the right-side window and mirror, while the left screen displays the left-side window and mirror. The simulator records driving data at 60 Hz.

EEG data was collected using Muse^{TM} by Interaxon (Fig. 1), a wireless nonintrusive headband consisting of 2 dry sensors located at Fp1 (frontal polar left) and Fp2 (frontal polar right) positions and two gel foam electrodes at TP9 and TP10 positions. The EEG headband was worn around the forehead (Fp1 and Fp2) with two electrodes attached behind the ears (TP9 and TP10). However, due to poor signal quality at TP9 and TP10, only the signals from Fp1 and Fp2 were analyzed. Thus, the results that we report are obtained using only dry electrodes. Fp1 and Fp2 channels are commonly used in consumer-grade EEG systems as they are not covered by hair and are easier to access with a simple band. These channels are in the frontal region, which has been found to show reactivity to workload changes as discussed earlier. The MuseLab software was used to record and analyze the EEG signals; the sampling frequency was 220 Hz and the software calculated the power of EEG bands at 10 Hz.

ECG, GSR, and respiration sensors by Becker Meditec, widely used in previous research [e.g., 35, 36], collected data at 240 Hz using the D-Lab software developed by Ergoneers. Solid gel foam electrodes were used for the ECG and GSR sensors (Fig. 1). ECG was recorded with three electrodes placed on participant's chest. The GSR sensors were attached beneath the bare left foot with one sensor in the middle and the other under the heel. The respiration band (Fig. 1) was worn around the chest or abdomen, at the position that exhibited most heaving when participants breathed. Gaze information was collected at 60 Hz through faceLAB^{TM} 5.0, a dashboard mounted eye-tracker by Seeing Machines (Fig. 1).

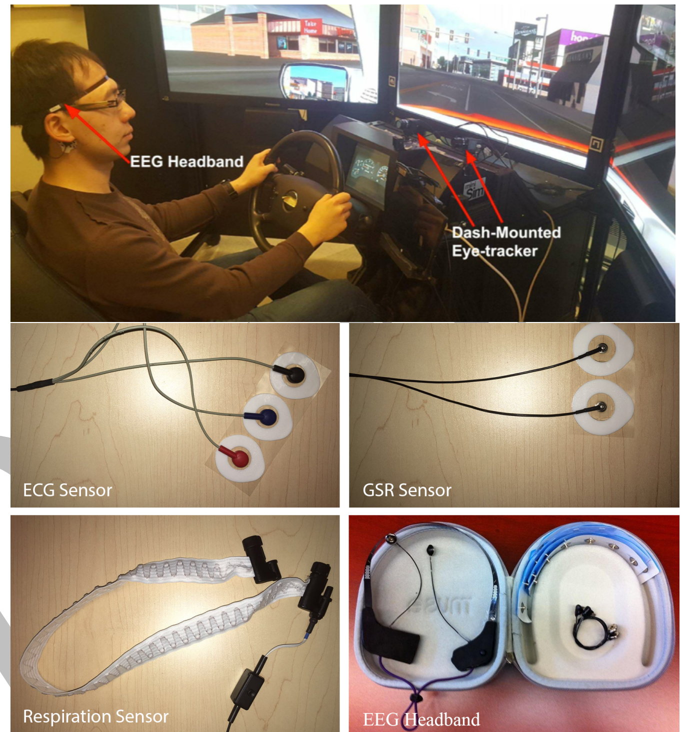


Fig. 1. Equipment: NADS miniSim^{TM} driving simulator, EEG headband, faceLAB^{TM} eye-tracker, and ECG, GSR, and respiration sensors.

C. Modified N-Back Task

A modified version of the n-back task based on the one utilized in [9] and [11] was used to introduce external cognitive load. The original n-back task used in [9] and [11] requires participants to listen to a series of single-digit numbers and respond verbally with the digit that was presented n-positions before (n-back) the current number, right after the current number is read to them. Therefore, there is continual verbal response during auditory stimulus presentation. Considering that facial muscle movements can interfere with EEG signals, and that such an artifact of the task paradigm should be avoided, we developed a modified version of this n-back task. Participants listened to a pre-recorded series of 10 letters, separated by approximately 2.5 second intervals, for an overall duration of approximately 25 seconds for each n-back task; these durations were in line with [9] and [11]. For the 1-back task, which was expected to impose less cognitive load than the 2-back task, participants were asked to count the number of times two identical letters appeared in pairs in a sequence, e.g., 'FF'. For the 2-back task, participants were asked to count the number of times two identical letters appeared in pairs with one letter in between, e.g., 'BHB'. Instead of answering during stimulus presentation, participants were asked to verbally respond with the total count of n-back instances at the end of each series. Letters instead of numbers were used in the modification to minimize the interference in working memory of the running total (of n-back instances in a series) with the auditory stimulus. Given the larger memory requirement, the modified n-back task was hypothesized to be more difficult than the n-back task used in [9] and [11], but still be able to maintain the order of difficulty from the 1-back to the 2-back level.

D. Driving Task

The driving scenarios required the participants to follow a lead vehicle at a speed of 64.4 km/h (40 mph) on a 4-lane urban route with light ambient traffic and some vehicles parked on the sides. In each n-back drive, the participants were presented with two groups of n-back tasks, each on a straight section of the route. Each group consisted of three n-back tasks (a series of 10 letters each), totalling to six n-back tasks completed within each drive. A notification and a brief reminder of the task was provided before each group to let the participant know that the n-back task was starting. At the end of each group, another notification was provided to let the participant know that the task had ended.

To simulate realistic driving scenarios and to gather reaction times to roadway events, the lead vehicle braked once per n-back task group (deceleration of 6 m/s^2), resulting in two braking events experienced during the n-back task within a drive. These braking events happened randomly during either the first or the third n-back task within a group. Two corresponding braking events occurred in the baseline drive and were positioned in the same section of the route where the n-back tasks were presented. Prior to the braking events, the lead vehicle speed was adjusted to create a 2 s headway time between the participant and the lead vehicles. The headway times achieved at the lead vehicle brake onset varied due to vehicle dynamics (mean=2.11 s, SD=0.56 s).

E. Procedures

Participant eligibility was verified and consent was obtained upon arrival. Participants first went through a practice drive in the simulator, on a route identical to the one used in the experimental drives. They practiced following the lead vehicle at a 2 second headway time and experienced lead vehicle braking events as they would happen in the experimental drives. They were then given written and oral instructions on the modified n-back task and practiced it without driving to ensure that they fully understood and were capable of doing the task. Physiological sensors were then placed on participants and the eye tracker was calibrated.

Next, participants completed another practice drive, this time performing the n-back task. However, they were told that this was an experimental drive in order to minimize their anticipation of where and when lead vehicle braking events were to occur in the actual experimental drives. The course was the same as the experimental drives and the earlier practice drive. In this drive, participants were given a group of three 1-back tasks and a group of three 2-back tasks. Multiple braking events were presented in each group of tasks to minimize participants' anticipation of the systematic nature of braking events that were going to happen in the experimental drives. Participants were also introduced to the NASA-TLX questionnaire at the end of this practice drive. The participants then completed the three experimental drives. NASA-TLX was collected after each drive through an online survey. Participants were given a 5-minute break after each drive. At the end of the experiment, participants were debriefed and received their payment.

F. Dependent Variables

As mentioned earlier, stimulus presentation for a given n-back task was approximately 25 seconds long. The data from these road segments were used in the analysis of n-back task effects; n-back task segments that had a lead vehicle braking event were only used for assessing lead vehicle braking response and n-back task performance, and were excluded from all other analysis. Data collected on the corresponding road segments of the baseline drive for approximately equal duration was used for comparison purposes.

Power spectrum density (PSD), $S_x(f)$, which describes the distribution of power into frequency components composing a signal, was calculated using Fast Fourier Transformation (FFT, [37]) method with a hamming window of 256 samples and overlap of 234 samples. Then, the power of alpha and theta EEG bands (alpha: 7.5-13 Hz; theta: 4-8 Hz) for each channel (Fp1 and Fp2) were calculated using integration:

$$P = \int_{f_1}^{f_2} S_x(f)df$$

where, P refers to the power of an EEG band; $S_x(f)$ is the PSD of the EEG signal; and f_1 and f_2 are the lower and upper range of the frequency range (e.g., 7.5 to 13 Hz for the alpha band). The power of each band was then averaged over each 25-second segment, which resulted in 4 (an alpha and a theta power value for each channel and 2 channels total) EEG data points for each 25-second segment. In layman's terms, the power of a band (e.g., power of alpha band) can be considered as the amount of activity found in a signal within a particular frequency range.

The experimenters logged the n-back responses manually and the percent correct rate was calculated after data collection. This rate was calculated by dividing the number of correct responses by six (number of n-back tasks within each drive). As mentioned previously, out of the six n-back tasks experienced within a drive, two had a lead vehicle braking event. Considering that a braking response might influence n-back performance, a second rate was also calculated for the four n-back tasks that did not correspond to a lead vehicle braking event.

Heartbeat identification was performed in MATLAB, using the signal processing toolbox. A moving average method with a window size of 1/6 seconds was adopted to remove the noise in the respiration data. HR, respiration rate, and blink rate were calculated as frequency over each 25-second segment, whereas GSR, respiration depth, average diameter of the left and right pupils, and vehicle speed were calculated as averages over each 25-second segment. SD of gaze position and SD of vehicle speed were also obtained over each 25-second segment. Accelerator release time (ART) was calculated from the lead vehicle brake light onset to the participant's foot fully releasing the brake pedal (SAE J2944_201506). The calculation of NASA-TLX scores followed the method outlined in [38].

G. Data Analysis

Secondary task performance was analyzed using the Friedman test. Other analyses were conducted through mixed linear models, with cognitive taskload as a fixed and participant as a random factor. Mixed linear models were built in PROC MIXED in SAS. Variance-covariance structures were selected based on the Bayesian Information Criterion. Normality and homoscedasticity were checked. In the mixed linear models, from each drive, four data points (from four data segments without lead vehicle braking) were used for EEG, HR, GSR, respiration rate, blink rate, SD of gaze position, and average pupil diameter; two data points were used for ART (from two data segments with lead vehicle braking); and one data point was used for NASA-TLX (collected at the end of each drive).

IV. EXPERIMENTAL RESULTS

Because of technical and data quality issues, GSR data from 3 participants (due to sensor detachment), HR data from 4 participants (due to sensor detachment), and ART data from 1 participant were lost (due to simulator software malfunction). Respiration and pupil diameter data were particularly noisy and thus were unidentifiable at times, leading to the removal of 61 data points for respiration rate, 73 data points for respiration depth, and 150 data points for pupil diameter (out of 444 total data points for each). Although headway time was controlled in the experiment to minimize variance in how participants experienced lead vehicle braking events, it was not possible to perfectly control headway due to vehicle dynamics. There were five data points where the participants failed to properly follow the lead vehicle, resulting in ARTs that were particularly long (studentized residual >4). Thus, these five data points were also removed from analysis. Headway time was included as a covariate in the analysis of ART.

The results presented in the following section, in Fig. 2 and Table 2, serve as validation for our modified n-back task. EEG results are presented in Fig. 3 and Table 3.

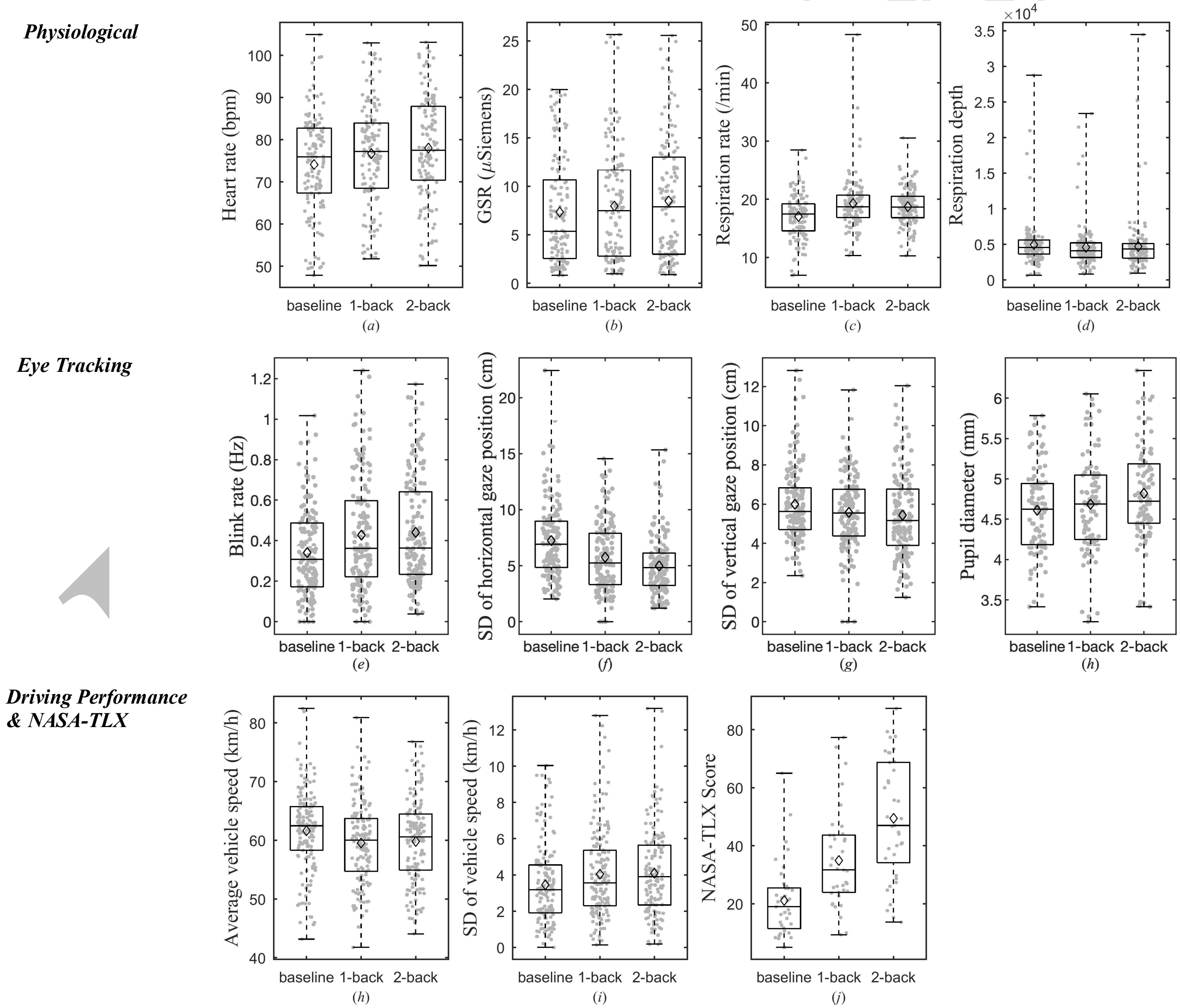


Fig. 2. Boxplots for physiological, eye tracking, driving performance, and subjective measures. The boxplots present minimum, first quartile, median, third quartile, and maximum, as well as data points indicated with gray circles and means indicated with hollow diamonds.

TABLE 2
VALIDATION OF THE N-BACK TASK MODIFICATION: ESTIMATED DIFFERENCE (Δ), 95% CONFIDENCE INTERVALS (CI), AND P-VALUES FOR PAIRWISE COMPARISONS

	1-back vs. Baseline			2-back vs. 1-back			2-back vs. Baseline		
Measures	*Δ*	*95% CI*	*p-value*	*Δ*	*95% CI*	*p-value*	*Δ*	*95% CI*	*p-value*
Heart rate (beats per min, bpm)	2.5	1.6, 3.5	<.0001	1.3	0.4, 2.2	.006	3.9	2.9, 4.8	<.0001
GSR (μSiemens)	1.1	0.6, 1.6	<.0001	1.0	0.6, 1.5	<.0001	2.1	2, 2.8	<.0001
Respiration									
Rate (per minute)	1.9	1.3, 2.5	<.0001	N.S.			1.7	1.1, 2.3	<.0001
Depth	-282.4	-442.8, -122.0	.0008	N.S.			-229.0	-338.0, 70.0	.006
Eye tracking									
Gaze position SD (cm)									
horizontal	-2.0	-2.5, -1.5	<.0001	-0.8	-1.2, -0.2	.002	-2.8	-3.2, -2.3	<.0001
vertical	-0.7	-1.1, -0.2	.007	N.S.			-1.0	-1.6, -0.4	.001
Blink rate (Hz)	0.07	0.03, 0.10	.0003	N.S.			0.08	0.05, 0.12	<.0001
Pupil diameter (mm)	0.11	0.05, 0.18	.0006	0.12	0.06, 0.18	.0004	0.23	0.17, 0.30	<.0001
Driving Performance									
Average speed (km/h)	-2.1	-3.1, -1.0	.0002	N.S.			-1.6	-2.8, -0.8	.0009
SD of speed (km/h)	0.59	0.12, 1.05	.01	N.S.			0.23	0.18, 1.11	.008
NASA-TLX	13.8	8.0, 19.7	<.0001	14.5	8.6, 20.3	<.0001	28.3	22.5, 34.2	<.0001

N.S.: NON-SIGNIFICANT (P > .05)

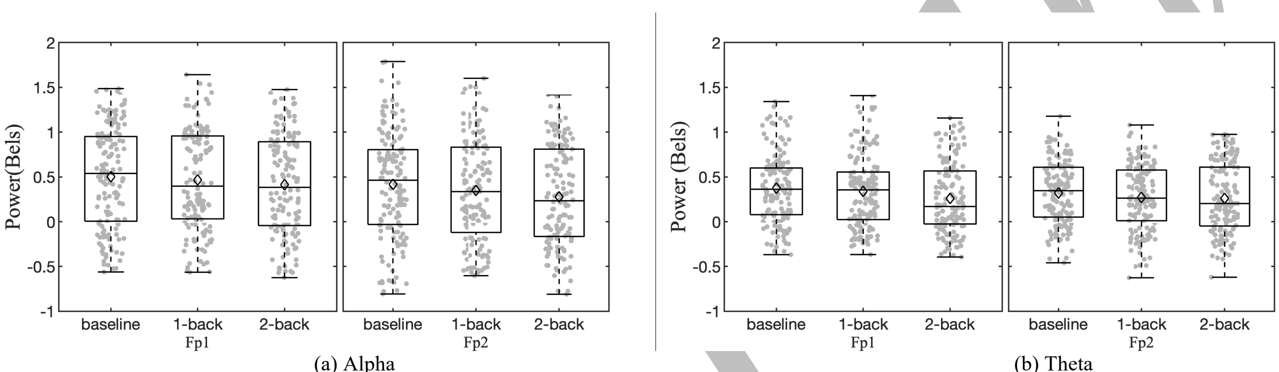


Fig. 3. Boxplots for EEG results: the power of (a) alpha and (b) theta bands at Fp1 and Fp2 positions.

TABLE 3
PAIRWISE COMPARISONS FOR POWER OF EEG BANDS

	1-back vs. Baseline			2-back vs. 1-back			2-back vs. Baseline		
EEG Measures	*Δ*	*95% CI*	*p-value*	*Δ*	*95% CI*	*p-value*	*Δ*	*95% CI*	*p-value*
Alpha (Bels)									
Fp1	-0.092	-0.152, -0.031	.003	-0.117	-0.178, -0.057	.0002	-0.209	-0.293, -0.125	<.0001
Fp2	-0.179	-0.256, -0.103	<.0001	-0.147	-0.224, -0.070	.0003	-0.326	-0.432, -0.221	<.0001
Theta (Bels)									
Fp1	N.S.			-0.077	-0.122, -0.031	.001	-0.113	-0.159, -0.068	<.0001
Fp2	-0.083	-0.142, -0.026	.007	-0.076	-0.136, -0.017	.01	-0.159	-0.240, -0.078	.0002

N.S.: NON-SIGNIFICANT (P > .05)

A. *Validation of the Modified N-Back Task*

The results reported in this section show that the modified n-back task influenced secondary task performance, ECG, GSR, respiration, eye tracking, driving performance, and subjective workload as expected based on previous research.

As mentioned earlier, a total of six n-back tasks were completed for each n-back drive. Correct response rate for 2-back (mean: 66.7%, SD: 21.2) was lower than 1-back (mean: 94.1%, SD: 9.8%), $\chi^2(1) = 31.0$, $p < .0001$. When the n-back tasks that corresponded to a lead vehicle braking event (2 per drive) were excluded from analysis, the 2-back task (mean: 71.6%, SD: 25.8) still had a lower correct response rate than the 1-back task (mean: 95.9%, SD: 9.3), $\chi^2(1) = 19.2$, $p < .0001$.

Heart rate ($F(2, 66) = 35.8$, $p < .0001$), GSR ($F(2, 64) = 21.48$, $p < .0001$), SD of horizontal gaze position ($F(2, 72) = 75.04$, $p < .0001$), average pupil diameter ($F(2, 47) = 28.66$, $p <.0001$), and NASA TLX ($F(2, 72) = 46.34$, $p < .0001$) showed significant stepwise trends with increasing cognitive load. Heart rate increased by 2.5 bpm from baseline to 1-back, and by 1.3 bmp from 1-back to 2-back. GSR increased by 1.1 μSiemens from baseline to 1-back, and by 1.0 μSiemens from 1-back to 2-back. SD of horizontal gaze position decreased by 2.0 cm from baseline to 1-back, and by 0.8 cm from 1-back to 2-back. The average pupil diameter increased by 0.11 mm from baseline to 1-back and by 0.12 mm from 1-back to 2-back.

Finally, NASA TLX increased by 13.8 from baseline to 1-back, and by 14.5 from 1-back to 2-back.

Respiration rate ($F(2, 63) = 25.83$, $p < .0001$) and depth ($F(2,61) = 6.93$, $p = 0.002$), SD of vertical gaze position ($F(2, 72) = 5.98$, $p = .004$), blink rate ($F(2, 72) = 11.89$, $p < .0001$), average speed ($F(2, 72) = 9.29$, $p = .0003$), and SD of speed ($F(2,72) = 4.61$, $p = .01$) showed a response to added cognitive load, but with no significant differences between 1-back and 2-back levels. ART was found to increase with increasing headway time ($F(1, 151) = 9.69$, $p = .002$); cognitive taskload did not have an effect on ART ($F(2, 184) = 1.82$, $p = .17$).

B. Effects of the Modified N-Back Task on EEG

For EEG signals, both the power of alpha (Fp1: $F(2, 72) = 12.42$, $p < .0001$; Fp2: $F(2, 72) = 19.15$, $p < .0001$) and theta bands (Fp1: $F(2, 72) = 12.84$, $p < .0001$; Fp2: $F(2,72) = 7.71$, $p = .0009$) decreased stepwise with increasing cognitive load, except for a non-significant comparison between baseline and 1-back for Fp1 theta band (Fig. 3, Table 3).

V. Discussion

In a driving simulator study with 37 participants, we explored the influence of external cognitive taskload on driver's EEG signals collected through a consumer-grade EEG system at two frontal positions. The aim was to assess whether it is feasible to utilize a relatively non-intrusive and cheap EEG system to differentiate between different levels of cognitive taskload experienced by drivers; our study was the first to investigate this topic. Improvements in cost and intrusiveness bring EEG systems closer to in-vehicle implementation; however, research is needed to test whether these newer systems are sensitive enough to capture different driver states. Different levels of cognitive taskload were imposed on the participants through the n-back task, a commonly used task for studying working memory capacity [22]. In order to ensure that the EEG signal quality was not degraded by verbal responses that were an artifact of the n-back task paradigm, we modified the n-back task to move the verbal response required during stimulus presentation to after a string of stimuli is presented. The effectiveness of this modified n-back task on imposing increasing levels of cognitive load was validated using a variety of physiological measures along with eye-tracking, driving performance, and subjective measures.

A. The Effectiveness of the Modified N-back Task

In driving research, the n-back task has been implemented and widely used as auditory stimulus and continual verbal response [9, 11]: participants listen to a series of single-digit numbers and respond verbally with the digit that was presented n-positions before the current number, right after the current number is read to them. Our modification moved the verbal response required during stimulus presentation to after a string of stimuli is presented; and replaced numbers with letters to minimize the interference in working memory of the running total of n-back instances (required response) with the auditory stimulus. Overall, the modified n-back task was effective in imposing differentiable levels of cognitive load. Participants performed worse on the 2-back compared to 1-back task. Subjective workload increased from the baseline to 1-back, and from 1-back to 2-back. The physiological, eye-tracking, and driving performance results, which were in line with previous research, also suggest that the modification was successful in imposing differentiable levels of cognitive load.

Heart rate and GSR showed a stepwise increase from baseline to 1-back, and then to 2-back; a finding in line with the results of both [9] and [11], with the exception that [9] did not show a significant difference between 1-back and 2-back levels for GSR. The difference may be attributed to [9] being an on-road study, whereas both our study as well as [11] being conducted in a simulator. Our respiration rate results were same as the findings of [9]; [11] did not report respiration data. Further, we also looked at respiration depth, which was not a measure used in earlier studies of similar nature. We found that the statistical significances in respiration depth follow the results of respiration rate.

As for eye tracking measures, we found a stepwise decrease in horizontal gaze position variability and a stepwise increase in pupil diameter with increasing levels of cognitive taskload; [11] also found similar trends in horizontal gaze position but no difference between 1-back and 2-back; [9] did not report on gaze data. These earlier n-back studies also did not report on vertical gaze position variability, pupil diameter, and blink rate; however, our results on these measures were in line with [3, 19, 20], which investigated driver cognitive load using different task paradigms.

The two previous n-back studies discussed above [9, 11] reported conflicting results for average speed and SD of speed. Our study does not support one over the other, but introduces additional conflict. In our study, average speed was found to decrease and SD of speed was found to increase with increased taskload. In [9], which is an on-road study, an increase of average speed was observed; whereas in [11] average speed was found to decrease. Thus, our study is in line with [11] in terms of average speed. However, we found SD of speed to increase whereas [11] found it to decrease with increased taskload. It appears that the data collection medium as well as the particular scenarios utilized in different experiments have a large impact on the speed maintenance, potentially masking the effects of cognitive taskload.

It should be noted that the modification used in our n-back version that required the maintenance of the running total in working memory may have resulted in our task to be more difficult than the earlier n-back task used in the driving domain [9, 11]. The correct response rates observed in our study (1-back: 94%, 2-back: 68%) were lower than the ones observed in [9] (1-back: 98%, 2-back: 88%) and [11] (1-back: 95%, 2-back: 85%). Our response accuracy calculations are different than the ones used in [9, 11] given the differences in tasks, thus, the response rates found in our study may not be directly comparable to these earlier studies. Still, further research is needed to compare our task to earlier versions of n-back. It should also be noted that in our version of the n-back task, there is a chance that multiple mistakes made during a single task sequence may offset each other. This limitation should be considered by researchers who may put more emphasis on the task performance accuracy than we did in the current study; our motivation for developing the task was to increase taskload without degrading EEG signal quality, which we have achieved. Another limitation of our research is that we have not yet studied how the modified n-back task compares to common tasks that drivers perform in their vehicles. A comparison is needed such as the one reported in [23].

B. *EEG in Distinguishing Driver Cognitive Taskload*

The results showed that the consumer-grade EEG headset that was utilized in our experiment was sensitive in differentiating between the different levels of cognitive taskload imposed on the drivers. The two positions that were utilized, i.e., Fp1 and Fp2, are in the frontal region of the brain. Based on the literature discussed in our background section, we focused on alpha and theta bands for these two positions, which are known to show reactivity to high cognitive load levels. We found that the power of alpha band decreased for both positions with increasing taskload from the baseline to 1-back, and then to 2-back. This suppression of the alpha band trend observed in the frontal region was in line with both basic studies conducted in computer settings [25, 26] as well as a driving simulator study that used a visual-manual n-back task [31].

However, we found the power of the theta band to decrease as well with increasing cognitive taskload. Previous literature which reported an effect on theta for the frontal region, suggests the opposite direction, with theta band power increasing with increasing load. It should be noted that these earlier studies either focused on a single-task paradigm (i.e., the conduct of n-back alone with no additional tasks) [26, 27, 29] or utilized a visual-manual n-back task along with the primary task of driving [31]. The dissimilarities in task characteristics may help explain this seeming contradiction between our findings and those of earlier studies. Another potential explanation is that our participants may have experienced a higher degree of stress than those in [31] given that our participants drove through a more complex driving environment and arguably performed a more difficult secondary task. An n-back study conducted by [39] showed the power of the frontal theta band to decrease with increasing stress. It is also possible that the particular EEG device we used had issues related to capturing this frequency range. Further research is needed to identify the reasons of this contradicting finding.

Future research should also validate our findings and test the reliability of the sensors we used with a larger sample size, under different driving conditions, in field trials, and using research-grade EEG systems. Although we showed that a consumer-grade EEG system can differentiate between increasing levels of cognitive taskload imposed on drivers, we only utilized three difficulty levels. The sensitivity threshold of these systems can be assessed by investigating smaller changes in taskload. Further, we only focused on cognitive load imposed through the auditory channel and verbal coding. This paradigm is only representative of some of the external demands that drivers experience. Future research should also investigate different types of secondary tasks, including other cognitive secondary tasks that have been validated in earlier research, and whether and how these EEG systems react in response. Finally, although we have shown that a consumer-grade EEG system can provide useful information about driver state, these systems are still sensitive to artifacts such as facial or body movements that would naturally occur while driving and thus need further development before they can be successfully implemented within vehicles.

Acknowledgment

We gratefully acknowledge Amirhossein Aghaei, Winnie Chen, Mahdi Marsousi, George Liu, and Nicole Wongsoo for their contributions. Earlier findings from a limited number of initial participants were reported in the Proceedings of the Transportation Research Board Annual Meeting, 2017.

References

[1] T. A. Dingus *et al.*, "Driver crash risk factors and prevalence evaluation using naturalistic driving data," *Proceedings of the National Academy of Sciences,* vol. 113, no. 10, pp. 2636-2641, 2016.

[2] F. G. Paas and J. J. Van Merriënboer, "Instructional control of cognitive load in the training of complex cognitive tasks," *Educational Psychology Review,* vol. 6, no. 4, pp. 351-371, 1994.

[3] M. A. Recarte and L. M. Nunes, "Effects of verbal and spatial-imagery tasks on eye fixations while driving," *Journal of Experimental Psychology: Applied,* vol. 6, no. 1, p. 31, 2000.

[4] J. L. Harbluk, Y. I. Noy, P. L. Trbovich, and M. Eizenman, "An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance," *Accident Analysis & Prevention,* vol. 39, no. 2, pp. 372-379, 2007.

[5] B. Reimer *et al.*, "A driving simulator study examining phone dialing with an iphone vs. a button style flip-phone," in *Proceeding of Human Factors and Ergonomics Society Annual Meeting*, 2012, pp. 2191-2195, Los Angeles, CA: Sage Publications.

[6] K. Young and M. Regan, "Driver distraction: a review of the literature," in *Distracted Driving*, I. J. Faulks, M. Hammer, M. Stevenson, J. Brown, A. Porter, and J. D. Irwin, Eds. Sydney, NSW, Australia: Australasian College of Road Safety, 2007, pp. 379-405.

[7] Y. Liang, M. L. Reyes, and J. D. Lee, "Real-time detection of driver cognitive distraction using support vector machines," *IEEE Trans. Intell. Transp. Syst.,* vol. 8, pp. 340-350, 2007.

[8] K. Ryu and R. Myung, "Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic," *International Journal of Industrial Ergonomics,* vol. 35, no. 11, pp. 991-1009, 2005.

[9] B. Mehler, B. Reimer, J. F. Coughlin, and J. a. Dusek, "Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers," *Transportation Research Record,* no. 2138, pp. 6-12, 2009.

[10] S. Miller, "Literature review workload measures," University of Iowa, Iowa City, IA, USA, ID: N01-006, Aug. 2001.

[11] B. Mehler, B. Reimer, and J. F. Coughlin, "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,* vol. 54, no. 3, pp. 396-412, 2012.

[12] A. J. Camm *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," *Circulation,* vol. 93, no. 5, pp. 1043-1065, 1996.

[13] A. H. Roscoe, "Assessing pilot workload. Why measure heart rate, HRV and respiration?," *Biological Psychology,* vol. 34, no. 2, pp. 259-287, 1992.

[14] G. Borghini, L. Astolfi, G. Vecchiato, D. Mattia, and F. Babiloni, "Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness," *Neuroscience and Biobehavioral Reviews,* vol. 44, pp. 58-75, 2014.

[15] S. Fazli *et al.*, "Enhanced performance by a hybrid NIRS–EEG brain computer interface," *Neuroimage,* vol. 59, no. 1, pp. 519-529, 2012.

[16] Y. K. Wang, T. P. Jung, and C. T. Lin, "EEG-based attention tracking during distracted driving," *IEEE Trans. Neural Syst. Rehabil. Eng.,* vol. 23, no. 6, pp. 1085-1094, 2015.

[17] D. L. Strayer, J. Turrill, J. M. Cooper, J. R. Coleman, N. Medeiros-Ward, and F. Biondi, "Assessing cognitive distraction in the automobile," *Human Factors,* vol. 57, no. 8, pp. 1300-1324, 2015.

[18] K. A. Brookhuis, G. de Vries, and D. de Waard, "The effects of mobile telephoning on driving performance," *Accident Analysis & Prevention,* vol. 23, no. 4, pp. 309-316, 1991.

[19] Y. Liang and J. D. Lee, "Combining cognitive and visual distraction: less than the sum of its parts," *Accident Analysis & Prevention,* vol. 42, no. 3, pp. 881-890, 2010.

[20] M. A. Recarte and L. M. Nunes, "Mental workload while driving: effects on visual search, discrimination, and decision making," *Journal of Experimental Psychology: Applied,* vol. 9, no. 2, p. 119, 2003.

[21] J. A. Urigüen and B. Garcia-Zapirain, "EEG artifact removal—state-of-the-art and guidelines," *Journal of Neural Engineering,* vol. 12, no. 3, p. 031001, 2015.

[22] A. M. Owen, K. M. McMillan, A. R. Laird, and E. Bullmore, "N‐back working memory paradigm: a meta‐analysis of normative functional neuroimaging studies," *Human Brain Mapping,* vol. 25, no. 1, pp. 46-59, 2005.

[23] B. Mehler, B. Reimer, J. Dobres, and J. F. Coughlin, "MIT/CSRC project on assessing the demands of voice based In-vehicle interfaces phase II experiment 3 -2015 Toyota Corolla (2015b)," Massachusetts Institute of Technology, Cambridge, MA, USA,, Tech. Rep. 2015-14, 2015.

[24] S. Wang, J. Gwizdka, and W. A. Chaovalitwongse, "Using wireless EEG signals to assess memory workload in the n-back task," *IEEE Trans. Human–Mach. Syst.,* vol. 46, no. 3, pp. 424-435, 2016.

[25] A. Gevins and M. E. Smith, "Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style," *Cerebral Cortex,* vol. 10, no. 9, pp. 829-839, 2000.

[26] A. Gevins *et al.*, "Monitoring working memory load during computer-based tasks with EEG pattern recognition methods," *Human Factors,* vol. 40, no. 1, pp. 79-91, 1998.

[27] A. Gevins, M. E. Smith, L. McEvoy, and D. Yu, "High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice," *Cerebral Cortex,* vol. 7, no. 4, pp. 374-385, 1997.

[28] Y. Chen and X. Huang, "Modulation of alpha and beta oscillations during an n-back task with varying temporal memory load," *Frontiers in Psychology,* vol. 6, 2016, Art. no. 2031.

[29] M.-P. Deiber *et al.*, "Distinction between perceptual and attentional processing in working memory tasks: a study of phase-locked and induced oscillatory brain dynamics," *Journal of Cognitive Neuroscience,* vol. 19, no. 1, pp. 158-172, 2007.

[30] M. E. Smith, A. Gevins, H. Brown, A. Karnik, and R. Du, "Monitoring task loading with multivariate EEG measures during complex forms of human-computer interaction," *Human Factors,* vol. 43, no. 3, pp. 366-380, 2001.

[31] S. Lei and M. Roetting, "Influence of task combination on EEG spectrum modulation for driver workload estimation," *Human Factors,* vol. 53, no. 2, pp. 168-179, 2011.

[32] Y. K. Wang, S. A. Chen, and C. T. Lin, "An EEG-based brain-computer interface for dual task driving detection," *Neurocomputing,* vol. 129, pp. 85-93, 2014.

[33] S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty," *Science,* vol. 150, no. 3700, pp. 1187-1188, 1965.

[34] C.-T. Lin *et al.*, "Wireless and wearable EEG system for evaluating driver vigilance," *IEEE Trans. Biomed. Circuits Syst.,* vol. 8, no. 2, pp. 165-176, 2014.

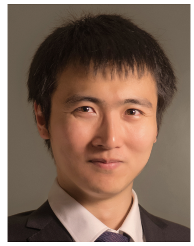
[35] M. Schultze-Kraft, S. Dähne, M. Gugler, G. Curio, and B. Blankertz, "Unsupervised classification of operator workload from brain signals," *Journal of Neural Engineering,* vol. 13, no. 3, p. 036008, 2016.

[36] A. Schwerdtfeger and P. Friedrich-Mai, "Social interaction moderates the relationship between depressive mood and heart rate variability: Evidence from an ambulatory monitoring study," *Health Psychology,* vol. 28, no. 4, p. 501, 2009.

[37] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.,* vol. 15, no. 2, pp. 70-73, 1967.

[38] S. Hart, G., "NASA-Task Load Index (NASA-TLX); 20 years later," in *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, 2006, pp. 904-908, Los Angeles, CA: Sage Publications.

[39] M. Gärtner, L. Rohde-Liebenau, S. Grimm, and M. Bajbouj, "Working memory-related frontal theta activity is decreased under acute stress," *Psychoneuroendocrinology,* vol. 43, pp. 105-113, 2014.



Dengbo He received his bachelor's degree in vehicle engineering from Hunan University, Changsha, China, in 2012 and his M.S. degree in mechanical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2016. He is currently a Ph.D. student at the University of Toronto in the Mechanical and Industrial Engineering Department. His research interests include human factors, driver behavior, and vehicle dynamics and control.

Birsen Donmez (M'10-SM'14) received her B.S. degree in mechanical engineering from Bogazici University, Istanbul, Turkey, in 2001, and M.S. and Ph.D. degrees in industrial engineering in 2004 and 2007, respectively, and M.S. degree in statistics in 2007 from the University of Iowa, Iowa City, IA, USA. She is currently an Associate Professor at the University of Toronto, Mechanical & Industrial Engineering and is the Canada Research Chair in Human Factors and Transportation.

Donmez serves on multiple Transportation Research Board committees and as an associate editor for IEEE Transactions on Human-Machine Systems and has served as the General Chair for AutomotiveUI'18.

Cheng Chen Liu received her B.A.Sc. and M.A.Sc. degrees in computer engineering from, University of Toronto, Toronto, Ontario, Canada in 2014 and 2017 respectively. Her research interests include machine learning and cognitive distraction in drivers.

Konstantinos N. Plataniotis (S'93-M'95-SM'03-F'12) is a Professor and the Bell Canada Chair in Multimedia with the ECE Department at the University of Toronto. He is the founder and inaugural Director-Research for the Identity, Privacy and Security Institute and has served as the Director for the Knowledge Media Design Institute at the University of Toronto from 2010 to 2012.

Dr. Plataniotis is a Fellow of the Engineering Institute of Canada. He has served as the Editor-in-Chief of the IEEE Signal Processing Letters, as General Co-chair for 2017 IEEE GlobalSIP and the 2018 IEEE International Conference on Image Processing, and as Technical Co-Chair of the IEEE 2013 International Conference in Acoustics, Speech and Signal Processing. He was the IEEE Signal Processing Society Vice President for Membership (2014 -2016).