

ASSESSING HIGH COGNITIVE LOAD IN DRIVERS THROUGH ELECTROENCEPHALOGRAPHY

Dengbo He

University of Toronto, Department of Mechanical and Industrial Engineering
5 King's College Road, Toronto, ON M5S 3G8, Canada
Tel: +1 647-995-4236 Fax: +1 416-978-7753 Email: dengbo.he@mail.utoronto.ca

Cheng Chen Liu

University of Toronto, The Edward S. Rogers Sr Department of Electrical and Computer Engineering
10 King's College Road, Toronto, ON M5S 3G4, Canada
Tel: +1 647-618-8186 Email: cheng.liu@mail.utoronto.ca

Birsen Donmez, Corresponding Author

University of Toronto, Department of Mechanical and Industrial Engineering
5 King's College Road, Toronto, ON M5S 3G8, Canada
Tel: +1 416-978-7399 Fax: +1 416-978-7753 Email: donmez@mie.utoronto.ca

Konstantinos Plataniotis

University of Toronto, The Edward S. Rogers Sr Department of Electrical and Computer Engineering
10 King's College Road, Toronto, ON M5S 3G4, Canada
Tel: +1 416-946-5605 Fax: +1 416-978-4425 Email: kostas@comm.utoronto.ca

Word count: 6,306 words text + 4 tables/figures x 250 words (each) = 7,306 words
Submission date: November 6, 2016

ABSTRACT

This paper explores the influence of high cognitive load on driver's Electroencephalography (EEG) signals collected from four positions (TP9, Fp1, Fp2, TP10) along with other physiological signals, plus eye tracking, driving performance, and subjective measures. Although EEG has been used in driving research to assess mental workload, only a few studies focused on high cognitive load, but they utilized research-grade EEG systems. Recent advancements allow for less intrusive and more affordable systems to be incorporated into vehicles. We tested the feasibility of one such system to differentiate three incremental levels of cognitive taskload in a preliminary simulator study, which so far has been completed by 15 participants. Each participant 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 during auditory stimulus presentation to increase EEG signal quality, with the 2-back level still imposing higher cognitive demand than 1-back. The system tested was sensitive to taskload levels, with alpha band being sensitive among all difficulty levels; beta and gamma bands distinguishing 2-back level from the baseline and 1-back; and the delta band distinguishing baseline from the n-back levels. In line with previous studies, galvanic skin response and standard deviation of gaze position also showed significant stepwise trends from the baseline to 1-back and then to 2-back. Further research is needed to investigate the ability of consumer-grade EEG headbands to differentiate different driver states.

Keywords: Driver Workload, Driver State, Electroencephalography, Physiological Measures, N-back Task

INTRODUCTION

Driving can be mentally demanding, especially under certain circumstances such as bad weather and complex traffic conditions. Activities secondary to driving, such as the use of in-vehicle infotainment systems and smart phones, can also claim mental resources. Heavy mental load can impair operator performance as can low levels of mental load (1). For instance, several simulator and on-road studies indicate that high levels of cognitive load impair drivers' visual scanning behaviors and driving performance (2, 3). Although drivers can moderate their cognitive load to some extent, such as by reducing their speed, avoiding lane changes, and increasing headway (4, 5), these acts may not be sufficient to fully compensate for the external demands placed on 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 mental load (e.g., (6)) and ensuing interventions.

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

TABLE 1 Summary of Example Mental Workload Measurements

Measure	Trend with Increased Cognitive Taskload
<i>Physiological</i>	
EEG	Power of alpha band ↓ (7, 8) P300 latency ↑ (9)
ECG	HR ↑ (10–12) HRV ↓ (12)
GSR	↑ (10, 11)
Respiration	Rate ↑ (10)
<i>Eye Tracking</i>	
Gaze position	Periphery, mirror, instrument check ↓ (2) SD of horizontal position ↓ (11, 13) SD of vertical position ↓ (13)
Eye blink	Frequency ↑ (13)
<i>Performance-Based</i>	
Vehicle speed	Average ↓ (11) SD ↑ (10) ↓ (11)
Steering wheel	Reversal rate ↑ (11)
<i>Subjective</i>	
NASA-TLX	↑ (2)

↑ increase; ↓ decrease

Previous studies have attempted to estimate drivers' mental workload using a variety of measures. In (10), external cognitive demand was introduced to the drivers in the simulator, through an auditory delayed digit recall task (14), an n-back task variation (15). This study utilized three n-back levels, 0-, 1-, and 2-back, corresponding to increasing levels of difficulty. Heart rate (HR) showed a stepwise increase with increasing n-back difficulty. GSR and

respiration rate showed significant increases from the baseline (no task) to the 0-back task and from the 0-back to 1-back, but no significant change was observed from 1-back to 2-back suggesting a plateau for these measures at higher levels of load tested. Although there appeared to be a decreasing trend for vehicle speed from the baseline up to the 1-back level, this trend was not significant, while there was a significant increase in speed from the 1-back to 2-back level. The standard deviation (SD) of speed also showed a significant increase from the 1-back to 2-back task. The same cognitive secondary task and task difficulty levels were used in an on-road study reported in (11). Both HR and GSR showed distinguishable increasing trends as task difficulty increased. Meanwhile, speed and the SD of speed showed a significant decrease with external cognitive demand; however, no differences were observed between the increasing levels of the n-back task. The steering wheel reversal rate was higher for 1- and 2-back levels compared with 0-back and baseline. The SD of horizontal gaze position decreased with added load but no difference was observed between 1- and 2-back levels. In both (10) and (11), the secondary task performance decreased with the increase of the n-back level, confirming the increasing difficulty associated with the three task levels utilized. Harbluk et al. (2) also conducted an on-road study that investigated the effects of cognitive demand on drivers. When drivers performed mental arithmetic tasks, they checked their periphery, mirrors, and instruments less frequently. In another on-road study, this time with a paced serial addition task on a hands-free telephone, HR was found to increase and HR variability was found to decrease (12). In (13), a driving simulator study, participants were presented with an auditory-spatial task that simulated extreme cognitive demand that drivers may experience while interacting with a navigation system. Blink frequency and SD of horizontal and vertical gaze position were found to increase with this task.

EEG is a measurement of the electrical activity of the brain (7). (16), (17), and (18) utilized EEG signals to classify driver state recorded in the simulator. Although all three studies imposed higher cognitive load through an external secondary task, all tasks were visual-manual in nature. Both (16) and (17) asked the participants whether a mathematical equation was correct or not and required a response through button presses. Although (16) does not explicitly state the presentation modality of this task, it appears to have utilized the same task as (17) given that the same researchers authored both publications. (18) used a visual-manual n-back task with responses collected through button presses. These studies did not exclude visual component from the secondary task aimed to impose high cognitive demand. In (9), an increased P300 peak latency was observed with higher cognitive demand in the laboratory setting 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 demand in the computer setting, but showed some sensitivity to added cognitive demand in the simulator. All of the above studies utilized complex EEG systems. Recent advancement in technology allows for less intrusive and more affordable EEG headbands, which, if able to detect mental load, can be used for real-time driver state detection in the car and accompanying interventions.

The correlation between EEG and high taskload has been widely studied and observed in non-driving domains, e.g., aviation (7). Although the frequency range varies across studies (7, 19), the power spectrum of the EEG signal is usually divided into five spectrums: delta, theta, alpha, beta, and gamma bands. Suppression of the alpha band was observed during complex and cognitively demanding tasks (7, 8). Increased EEG power spectra in the theta band is linked to a decrease in vigilance, while the increase of EEG power spectra in the beta band usually indicates increased alertness and arousal (7).

In this paper, we evaluate the feasibility of a consumer-grade EEG system to differentiate different levels of cognitive load in a preliminary driving simulator study with 15 participants. The relationship between the EEG signals collected from four positions (TP9, Fp1, Fp2, TP10) and varying levels of mental load imposed through a cognitive secondary task is investigated. The task used is a modified version of the n-back task. The n-back task has been used in a variety of EEG studies in non-driving domains (20–22); however, the tasks used in these studies involved visual stimulus presentation and manual response. Given that driving relies heavily on vision and manual control, we chose to use auditory stimulus presentation and verbal response as was done in (10, 11, 23). Furthermore, considering that EEG might be easily influenced by facial muscle movements, we modified the n-back task to remove the continual verbal response required from participants during auditory stimulus presentation.

Previous driving studies observed significant changes in HR, GSR, respiration rate, and gaze dispersion with increased difficulty (increased n-back level) in the auditory-verbal n-back task (10, 11). It is widely agreed that no single measure alone can provide sufficient information to estimate mental load (8, 10, 24). Further, given the modification applied to the n-back task in our study, other physiological, performance-based, eye-tracking, and subjective measures were thoroughly examined to ensure that the levels of the modified n-back task used in our study indeed increased cognitive load to levels that were distinguishable by at least some of these different measures.

METHODS

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

Participants

So far, 15 drivers (12 males and 3 females), recruited through campus and online posts, participated in this ongoing driving simulator study. Participants were required to drive at least several times per month, to hold a full driver's license (G license in Ontario, Canada or equivalent) for at least 3 years, and to be under 35 years old (average age: 27.6; SD: 4.45). To improve eye tracking quality, the participants were also required to be able to drive without glasses (contact lenses were allowed). 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 secondary task. The experiment took about 2.5 hours and all participants were paid the full bonus amount regardless of their performance.

Apparatus

The study was conducted on a NADS miniSim™ driving simulator (*Figure 1a*). This fixed-based simulator has three 42-inch screens, creating a 130° horizontal and 24° vertical field at a 48-inch 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. Driving data was recorded at 60 Hz. EEG data was collected using Muse™ by Interaxon (*Figure 1b*), a wireless noninvasive headband consisting of 2 dry sensors located at Fp1 and Fp2 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). The associated software, MuseLab, 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 collected data at 240 Hz using the D-Lab software developed by Ergoneers. Solid gel foam electrodes were used for the ECG (*Figure 1c*) and GSR sensors (*Figure 1d*). ECG was recorded with three electrodes, one placed on the neck over the vertebra, one placed on the left side of the ribcage over the second lowest rib, and one placed over the uppermost part of the center line of the ribcage. 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 (*Figure 1e*) was worn around the chest or abdomen, at the position that exhibited most heaving when the participants breathed. Gaze information was collected at 60 Hz through faceLABTM 5.0, a dashboard mounted eye-tracker by Seeing Machines.

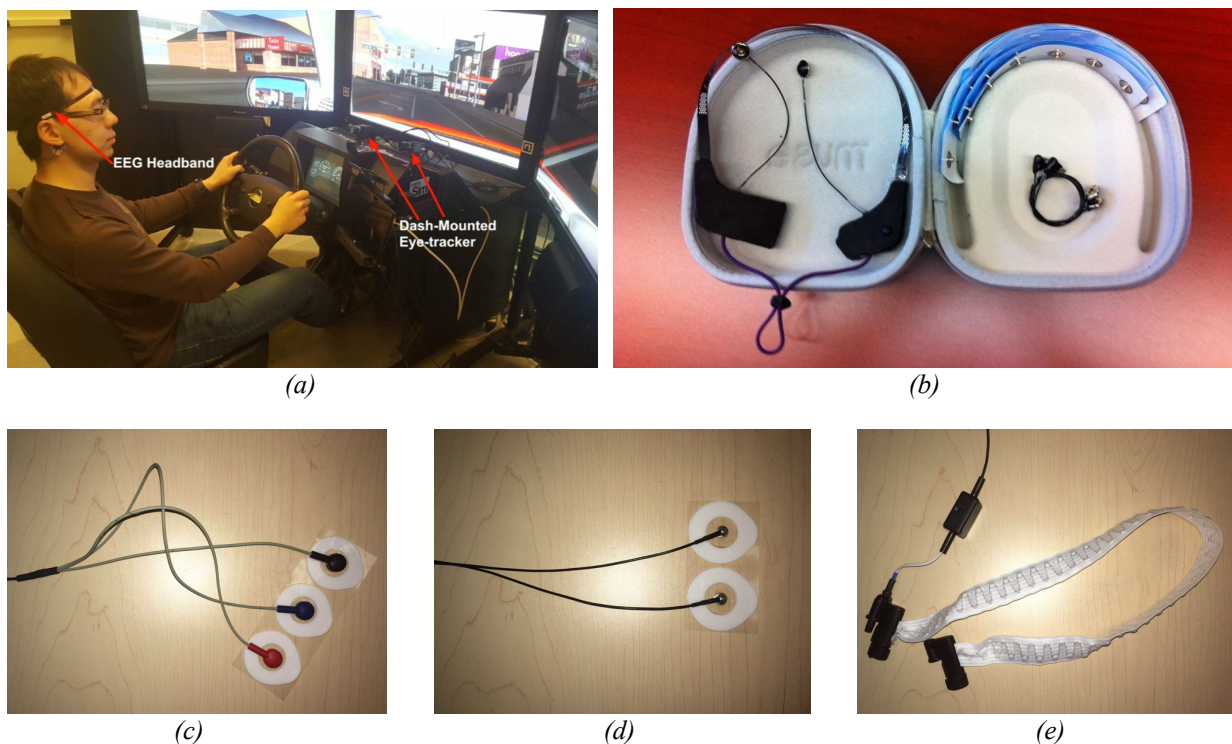


FIGURE 1 (a) NADS miniSimTM driving simulator. (b) MuseTM EEG headband. (c) ECG sensors. (d) GSR sensors. (e) Respiration band.

Secondary Cognitive Task

A modified version of the n-back task variation utilized in (10, 11) was used to introduce external cognitive load to participants. The original n-back task used in (10) and (11) required 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. Considering that facial muscle movements could interfere with the EEG signals, in this preliminary study, a modified version of the n-back task was used to remove verbal response during stimulus presentation. 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. 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., PP). 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., DTD). 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 is hypothesized to be more difficult than the n-back task used in (10) and (11), but still be able to maintain the order of difficulty from the 1-back to the 2-back level.

Driving Task

The driving scenarios required the participants to follow a lead vehicle at a speed of 40 mph on a 4-lane urban route with light ambient traffic and some vehicles parked on the sides. The scenarios were designed to involve mainly operational driving decisions, with no or minimal strategic or tactical decisions, such as navigation or passing a vehicle (25). The lead vehicle braked multiple times at a deceleration of 6 m/s^2 (intensive brake) or at 3 m/s^2 (slight brake). Prior to the braking events, the lead vehicle speed was adjusted to create a 2-sec time-headway between the participant and the lead vehicle. The gap times achieved at the lead vehicle brake onset varied due to vehicle dynamics (mean=2.11, SD=0.56).

In the n-back drives, the participants were presented with two groups of n-back tasks, each on a straight section of the route approximately 45 seconds apart. Each group consisted of three n-back tasks (a series of 10 letters each) presented consecutively, 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 n-back task, another notification was provided to let the participant know that the task had ended. There was one intensive lead vehicle braking event per group, resulting in two intensive 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. The two corresponding braking events in the baseline drive were positioned in the same section of the route where the n-back tasks were presented. In addition, two braking events (1 intensive brake and 1 slight brake) were introduced at random points before the first group of tasks were presented (and at the corresponding location for the baseline drive) in order to minimize participant anticipation of the braking events.

Procedures

Participant eligibility was verified and consent form was signed 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-sec gap time and experienced lead vehicle braking as it 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 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 a group of n-back 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.

Participants went on to complete the three formal experimental drives. NASA-TLX was collected after each drive through an online survey along with other questionnaires that are not reported in this paper. Participants were given a 5-minute break after each drive. At the end of the experiment, participants were briefed and received their payment.

Dependent Variables

The experimenters logged the n-back responses manually and the percent correct rate was calculated after data collection. This rate was calculated by dividing number of correct responses by six (n-back tasks). 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.

The two segments of data that involved a lead vehicle braking event were used to assess lead vehicle braking response, and the remaining four were used to assess all other measures that are detailed below. Each segment was approximately 25 seconds long, corresponding to the auditory stimuli presentation, excluding participants' verbal responses. Data collected on the corresponding segments of the baseline drive for approximately equal duration was used for comparison purposes.

For the EEG data, there were four signal channels corresponding to four positions of the sensors (TP9, Fp1, Fp2, TP10). MuseLab calculated the power of EEG bands (delta: 1-4Hz; theta: 4-8Hz; alpha: 7.5-13Hz; beta: 13-30Hz; gamma: 30-44Hz) for each channel using the Fast Fourier Transformation method with a Hamming window of 256 and overlap of 234 samples. The power bands were first averaged over each 25-second segment, and then were averaged across the four channels, which resulted in five EEG band power values for each 25-second segment. Heartbeat identification was performed in MATLAB signal processing toolbox. HR was calculated directly from heartbeat intervals. A moving average method with a window size of 1/6 seconds was adopted to remove the noise in the respiration data. Sixteen data points for respiration still contained excessive noise and were excluded from analysis.

Blink frequency and the horizontal and vertical gaze positions were obtained from the faceLABTM output. The horizontal and vertical gaze positions are the intersections of gaze vectors with the plane where the simulator's centre screen resides (approximately 48 inches away from the participant). From the simulator output, vehicle speed and the SD of vehicle speed, along with brake response time (BRT) were extracted as driving performance measures. Brake response time (BRT) was obtained from the lead vehicle brake light onset to the participant's foot contacting (start pressing) the brake pedal (SAE J2944_201506). The calculation of NASA-TLX scores followed the method outlined in (26).

RESULTS

Secondary task performance was analyzed using the Friedman test. Other statistical analysis was conducted through mixed linear models, with experimental condition as a fixed and participant as a random factor. Mixed linear models were built in PROC MIXED in SAS University Edition.

Variance-covariance structures were selected based on the Bayesian Information Criterion. Normality and homoscedasticity were checked and the data was transformed when necessary. In these mixed linear models, from each drive, four data points (from four data segments with no lead vehicle braking) were used for EEG, HR, GSR, and respiration rate, two data points were used for BRT (from two data segments with lead vehicle braking), and one data point was used for NASA-TLX (collected at the end of each drive).

TABLE 2 Significant ($p < .05$) Pairwise Comparisons from Mixed Linear Models

Signals	Baseline vs. 1-back Δ (95% CI)	1-back vs. 2-back Δ (95% CI)	Baseline vs. 2-back Δ (95% CI)
<i>Power of EEG bands</i>			
Alpha (Bels)	-0.070 (-0.126, -0.013)	-0.084 (-0.143, -0.027)	-0.154 (-0.231, -0.076)
Beta (Bels)	-0.059 (-0.120, 0.002) *	-0.062 (-0.123, -0.0008)	-0.121 (-0.204, -0.038)
Gamma (Bels)	-0.060 (-0.120, 0.0004) *	-0.063 (-0.123, -0.002)	-0.122 (-0.205, -0.040)
Delta (Bels)	-0.048 (-0.095, -0.001)	N.S.	-0.080 (-0.127, -0.034)
<i>ECG</i>			
Heart rate (bpm)	3.4 (1.3, 5.5)	N.S.	4.0 (1.1, 7.0)
GSR (μ Siemens)	1.6 (0.9, 2.2)	1.3 (0.7, 2.0)	2.9 (2.0, 3.8)
<i>Respiration</i>			
Rate (/min)	1.6 (0.9, 2.3)	N.S.	1.4 (0.8, 2.1)
<i>Eye tracking</i>			
Gaze position SD (cm)			
horizontal	-1.9 (-2.9, -0.9)	-1.4 (-2.4, -0.3)	-3.3 (-4.6, -1.9)
vertical	-0.7 (-1.2, -0.3)	N.S.	-1.0 (-1.4, -0.5)
<i>Driving Performance</i>			
Average velocity (mph)	-1.3 (-2.3, -0.4)	N.S.	-1.4 (-2.3, -0.4)
NASA-TLX	14.3 (4.4, 24.2)	18.1 (8.2, 28.0)	32.4 (22.5, 42.3)

* marginally significant ($.05 < p < .1$); N.S. non-significant ($p > .1$)

Secondary Task Performance

A total of six n-back tasks were completed for each n-back drive. Correct response rate for the 2-back task (mean: 65.6%, SD: 19.4%) was lower than the 1-back task (mean: 93.3%, SD: 10.5%), $\chi^2(1) = 13.0$, $p = .0003$. When the n-back tasks that corresponded to a lead vehicle braking event (2 per drive) were excluded from analysis, the correct response rates were similar, with the 2-back task (mean: 68.3%, SD: 20.0%) still leading to worse performance than the 1-back task (mean: 96.7%, SD: 8.8%), $\chi^2(1) = 11.0$, $p = .0009$.

EEG

The power of alpha ($F(2, 28) = 8.39$, $p = .001$), beta ($F(2, 28) = 4.43$, $p = .02$), gamma ($F(2, 28) = 4.65$, $p = .02$), and delta ($F(2, 28) = 6.25$, $p = .006$) bands were all significantly influenced by experimental condition (Figure 2). The trend for the power of alpha band was a decreasing one with increased cognitive load. There was a significant decrease from baseline to 1-back (0.070 Bels, $t(28) = 2.51$, $p = .02$), from baseline to 2-back (0.154 Bels, $t(28) = 4.08$, $p = .0003$) and from 1-back to 2-back (0.084 Bels, $t(28) = 3.03$, $p = .005$). For the power of beta and gamma bands, significant decreases were observed from baseline to 2-back (Beta: 0.121 Bels, $t(28) = 2.97$, $p = .006$; Gamma: 0.122 Bels, $t(28) = 3.05$, $p = .005$) and from 1-back to 2-back (Beta: 0.062 Bels, $t(28) = 2.07$, $p = .047$; Gamma: 0.063 Bels, $t(28) = 2.12$, $p = .043$), and there were marginally significant decreases from baseline to 1-back (Beta: 0.059 Bels, $t(28) = 1.98$, $p = .058$; Gamma: 0.060 Bels, $t(28) = 2.04$, $p = .051$). For the power of delta band, there was a

significant decrease from baseline to 1-back (0.048 Bels, $t(28) = 2.10$, $p = .045$) and to 2-back (0.080 Bels, $t(28) = 3.51$, $p = .002$).

ECG

Heart rate was significantly affected by experimental condition, $F(2, 28) = 5.72$, $p = .008$. The mean heart rate during the 1-back task was 3.4 beats per minute (bpm) higher than that of the baseline, $t(28) = 3.31$, $p = .003$, and the mean heart rate during the 2-back task was 4.0 bpm higher than that of the baseline, $t(28) = 2.82$, $p = .009$.

GSR

GSR was significantly influenced by experimental condition, $F(2, 28) = 20.17$, $p < .0001$. From baseline to 1-back, 1-back to 2-back, and baseline to 2-back, GSR increased by 1.6 ($t(28) = 4.78$, $p < .0001$), 1.3 ($t(28) = 4.11$, $p = .0003$), and 2.9 μ Siemens ($t(28) = 6.33$, $p < .0001$), respectively.

Respiration

Overall, respiration rate ($F(2, 25) = 13.44$, $p = .0001$) was significantly influenced by experimental condition. 1-back and 2-back tasks resulted in an increase of 1.6 ($t(25) = 4.74$, $p < .0001$) and 1.4 ($t(25) = 4.31$, $p = .0002$) respirations per minute respectively compared with the baseline. No significant difference was observed between 1-back and 2-back tasks ($p = .7$).

Eye Tracking

No significant effect was observed for blink frequency ($F(2, 28) = 1.31$, $p = .3$). The SD of gaze position was significantly affected by experimental condition (horizontal: $F(2, 28) = 12.96$, $p = .0001$; vertical: $F(2, 28) = 9.04$, $p = .0009$). The SD of horizontal gaze position decreased by 1.9 cm from baseline to 1-back ($t(28) = 3.82$, $p = .0007$), by 1.4 cm from 1-back to 2-back ($t(28) = 2.70$, $p = .01$), and by 3.3 cm from baseline to 2-back ($t(28) = 5.04$, $p < .0001$). There was no difference between 1-back and 2-back for the SD of vertical gaze position. However, a decrease of 0.7 cm and 1.0 cm was observed from baseline to 1-back ($t(28) = 3.11$, $p = .004$) and 2-back ($t(28) = 4.07$, $p = .0004$), respectively.

Driving Performance

Vehicle speed was significantly affected by experimental condition, $F(2, 28) = 5.83$, $p = .008$. Both the 1-back (1.3 mph, $t(28) = 2.93$, $p = .007$) and the 2-back tasks (1.4 mph, $t(28) = 2.99$, $p = .006$) resulted in a decreased speed compared to the baseline. However, there was no significant difference between 1-back and 2-back tasks ($p = .95$). No effect was observed for the SD of vehicle speed ($p = .15$), nor the BRT ($p = .16$).

Subjective Response

There was a significant main effect of experimental condition on NASA-TLX, $F(2, 28) = 22.51$, $p < .0001$. NASA TLX increased with added levels of cognitive load (baseline vs. 1-back: 14.3, $t(28) = 2.95$, $p = .006$; baseline vs. 2-back: 32.4 ($t(28) = 6.69$, $p < .0001$; 1-back vs. 2-back: 18.1, $t(28) = 3.74$, $p = .0008$).

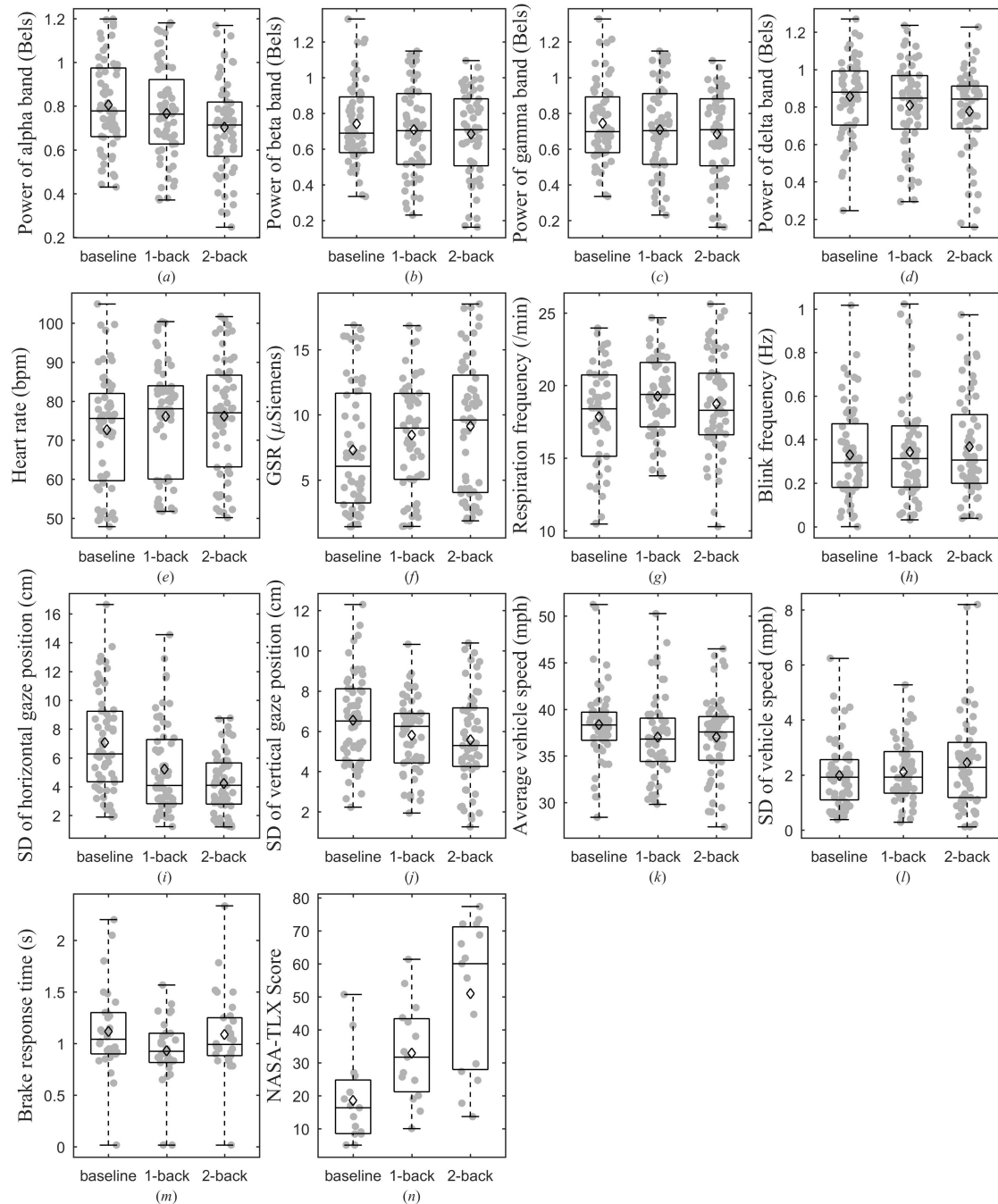


FIGURE 2 Boxplots presenting the minimum, 1st quartile, median, 3rd quartile, and maximum. The hollow dots represent the sample means whereas the gray dots represent data points. (a) Power of alpha band. (b) Power of beta band. (c) Power of gamma band. (d) Power of delta band. (e) Heart rate. (f) GSR. (g) Respiration frequency. (h) Blink frequency. (i) SD of horizontal gaze position. (j) SD of vertical gaze position. (k) Average vehicle speed. (l) SD of vehicle speed. (m) Brake response time. (n) NASA-TLX score.

DISCUSSION

In a preliminary driving simulator study with 15 participants, we explored the influence of external cognitive load on driver's EEG signals collected from four positions. Other physiological, eye-tracking, driving performance, and subjective measures were also analyzed.

Before discussing other results, the effectiveness of the newly introduced delayed-verbal-response n-back task must be examined. This modified n-back task was introduced to remove verbal responses during auditory stimulus presentation to increase EEG signal quality; however, it likely also increased the overall task difficulty compared to earlier auditory-verbal n-back tasks used in the driving domain (*10*, *11*). In fact, the correct response rates observed in our study (1-back: 93%, 2-back: 66%) were lower than the ones observed in (*10*) (1-back: 98%, 2-back: 88%) and (*11*) (1-back: 95%, 2-back: 85%). Our participants needed to keep in working memory not only letters but also an additional number for the running total of n-back instances presented to them. Therefore, there exists a concern that the 1-back level task might have already exhausted our participants' cognitive resources, and moving to 2-back could not have further increased their mental workload. The NASA-TLX scores showed a clear increasing trend with increasing n-back levels, indicating higher workload perceived by the participants. Furthermore, the 1-back task led to a higher correct response rate (93%) compared to the 2-back (66%). Considering the high correct rate observed in the 1-back task, it appears that working memory load was not exhausted at the 1-back level, with the 2-back level imposing higher cognitive demand than the 1-back, and the participants conducting the task at the 2-back level without entirely giving up.

Amongst all measures collected, three of them showed significant changes in all pairwise comparisons between experimental conditions (baseline, 1-back, and 2-back). The stepwise increase of GSR was consistent with (*11*), the on-road study, which also utilized an n-back task. Gaze dispersion assessed on the horizontal axis also showed the same trend as GSR, although no stepwise trend was found in (*11*) with no significant difference between 1-back and 2-back levels. In the driving simulator environment, our participants might not have had similar concerns of safety as the participants who drove in real traffic in (*11*). Another explanation is the higher difficulty experienced in our 2-back task compared to the one used in this earlier on-road study. From the newly introduced measurement, EEG, the alpha band demonstrated comparable performance with these two established measures (GSR and SD of horizontal gaze position). In other domains, the alpha band suppression has been linked to high mental workload (*7*). Hence, our results in driving are in line with other domains.

Respiration rate, HR, and delta band of EEG were only able to identify whether or not the secondary task was present, but not its level of difficulty. This finding was anticipated in the case of respiration rate, as (*10*), which also utilized the n-back task in the simulator, did not find a difference between 1- and 2-back levels. However, in contrast with (*10*) and (*11*), HR plateaued in our study at the 1-back level, suggesting that a limit of task difficulty might have been reached for this measure with our modified task, which was harder than the version used in (*10*) and (*11*) both for the 1-back and 2-back levels. The power of delta band showed a significant decrease from the baseline to the 1-back and 2-back task levels. Previous research only showed a link between increased delta band power with transition to a mental fatigue state (*7*). Thus, the decrease of the delta band power observed with added workload in our experiment needs further research.

Several other measures were found to be sensitive at the higher cognitive demand level, but not at the lower. The power of beta and gamma bands distinguished the 2-back level from the baseline and 1-back levels, and revealed a marginally significant difference between the 1-back

and 2-back levels. Because the modified 2-back task was in general quite difficult (66% correct response), it is reasonable to suspect that in addition to increased working memory load, the drivers may have experienced stress, task overload, and even mental fatigue. In fact, beta suppression has been linked to mental fatigue (7). The potential ability of the different EEG bands to capture high cognitive load along with task overload and fatigue states are worth examining in the future. If confirmed, this ability would place EEG in a unique position among other measures to identify cognitively overloaded driving states.

Finally, several measures did not show any significant changes with added cognitive demand. The lack of significance can be due to the limited sample size of our experiment. However, it can also be because some measures are not affected by cognitive load or are not as sensitive as other measures. For example, BRT was not significant in neither our study nor (13). Further, SD of vehicle speed showed contradictory trends between (10) and (11) as summarized in Table 1. Our study did not find a significant effect for this measure. The distortion of the vehicle speed and distance in the driving simulator (as reported by several participants) may be the reason. In contrast with (13), we did not find a significant effect of cognitive demand on blink rate. (13) used the same eye-tracking system as we did, but a different cognitive secondary task, which may explain the difference in our findings. Further, eye closure estimation was not always reliable in our system. Examining this variable through Electrooculography (EOG) sensors rather than eye-tracking can potentially provide increased accuracy.

In summary, this preliminary study explores the feasibility of using EEG signals collected by a consumer-grade headband for measuring mental workload during driving. Such applications are common in other domains, but have not yet been fully exploited for driver monitoring. To increase EEG signal quality, a modified n-back task was introduced, and its effectiveness on inducing incremental levels of mental workload was validated by various measures used in previous driving studies. Four bands, delta, alpha, beta, and gamma, were significantly influenced by increased cognitive taskload, with beta and gamma being sensitive at higher task difficulty, delta being significant to the presence of added cognitive demand, and alpha being the most sensitive by distinguishing all cognitive demand levels used in this study. Overall, EEG alpha band, GSR, and SD of horizontal gaze position appear to be the most sensitive to differentiating incremental levels of external cognitive load.

This is an ongoing experiment and 36 participants (18 males and 18 females) are planned. Two major limitations of this paper are the small sample size and unbalanced gender ratio (12 males and 3 females) of the data used in analysis. The potential distortion of speed and distance in the simulator as well as the lack of actual crash risk can also be a problem. Future research should validate our findings and test the reliability of the sensors used with a larger sample size, under different driving conditions, and in field trials.

ACKNOWLEDGEMENTS

The funding for this study was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC). We gratefully acknowledge Amirhossein S. Aghaei, Huei-Yen Winnie Chen, Mahdi Marsousi, and George Liu for their advice in experimental design and manuscript preparation, and Nicole Wongsoo for her help in data collection and cleaning.

REFERENCES

1. Proctor, R. W., and T. Van Zandt. *Human Factors in Simple and Complex Systems*. CRC Press, Boca Raton, 2008.
2. Harbluk, J. L., 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 and Prevention*, Vol. 39, No. 2, 2007, pp. 372–379.
3. Recarte, M. A., 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, 2000, pp. 31–43.
4. K. Young, and M. Regan. Driver Distraction : A Review of the Literature. In *Distracted Driving* (I.J. Faulks, M. Regan, M. Stevenson, J. Brown, A. Porter & J.D. Irwin eds.), NSW: Australasian College of Road Safety, Sydney, 2007, pp. 379–405.
5. Reimer, B., B. Donmez, M. Lavallière, B. Mehler, J. F. Coughlin, and N. Teasdale. Impact of Age and Cognitive Demand on Lane Choice and Changing under Actual Highway Conditions. *Accident Analysis and Prevention*, Vol. 52, 2013, pp. 125–132.
6. Liang, Y., M. L. Reyes, and J. D. Lee. Real-time Detection of Driver Cognitive Distraction Using Support Vector Machines. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 2, 2007, pp. 340–350.
7. Borghini, G., 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, 2014, pp. 58–75.
8. Ryu, K., 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, 2005, pp. 991–1009.
9. Strayer, D. L., J. Turrill, J. M. Cooper, J. R. Coleman, N. Medeiros-Ward, and F. Biondi. Assessing Cognitive Distraction in the Automobile. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 57, No. 8, 2015, pp. 1300–1324.
10. Mehler, B., B. Reimer, J. Coughlin, and J. Dusek. Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers. In *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2138, 2009, pp. 6–12.
11. Mehler, B., 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: The Journal of the Human Factors and Ergonomics Society*, Vol. 54, No. 3, 2012, pp. 396–412.
12. Brookhuis, K. A., G. de Vries, and D. de Waard. The Effects Of Mobile Telephoning On Driving Performance. *Accident Analysis and Prevention*, Vol. 23, No. 4, 1991, pp. 309–316.
13. Liang, Y., and J. D. Lee. Combining Cognitive and Visual Distraction: Less than the Sum of its Parts. *Accident Analysis and Prevention*, Vol. 42, No. 3, 2010, pp. 881–890.
14. Mehler, B., B. Reimer, and J. A. Dusek. *MIT AgeLab Delayed Digit Recall Task (n-back)*. MIT AgeLab white paper 2011-3B. Cambridge, Massachusetts, 2011.

15. Owen, A. M., 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, 2005, pp. 46–59.
16. Wang, Y. K., S. A. Chen, and C. T. Lin. An EEG-Based Brain-Computer Interface for Dual Task Driving Detection. *Neurocomputing*, Vol. 129, 2014, pp. 85–93.
17. Wang, Y. K., T. P. Jung, and C. T. Lin. EEG-Based Attention Tracking During Distracted Driving. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, Vol. 23, No. 6, 2015, pp. 1085–1094.
18. Lei, S., and M. Roetting. Influence of Task Combination on EEG Spectrum Modulation for Driver Workload Estimation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, Vol. 53, No. 2, 2011, pp. 168–179.
19. Lundqvist, M., P. Herman, and A. Lansner. Theta and Gamma Power Increases and Alpha/Beta Power Decreases with Memory Load in an Attractor Network Model. *Journal of Cognitive Neuroscience*, Vol. 23, No. 10, 2011, pp. 3008–3020.
20. Wang, S., J. Gwizdka, and W. A. Chaovalitwongse. Using Wireless EEG Signals to Assess Memory Workload in the N-Back Task. *IEEE Transactions on Human-Machine Systems*, Vol. 46, No. 3, 2016, pp. 424–435.
21. Chen, Y., 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, pp. 1–10.
22. Imperatori, C., B. Farina, R. Brunetti, V. Gnani, E. Testani, M. I. Quintiliani, C. Del Gatto, A. Indraco, A. Contardi, A. M. Speranza, and G. Della Marca. Modifications of EEG Power Spectra in Mesial Temporal Lobe During n-back Tasks of Increasing Difficulty. A sLORETA Study. *Frontiers in Human Neuroscience*, Vol. 7, 2013, pp. 109.
23. Zeitlin, L. R. Subsidiary Task Measures Of Driver Mental Workload: A Long-term Field Study. In *Transportation Research Record: Journal of the Transportation Research Board*, No. 1403, 1993, pp. 23-27
24. Dong, Y., Z. Hu, K. Uchimura, and N. Murayama. Driver Inattention Monitoring System for Intelligent Vehicles: A Review. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 12, No. 2, 2011, pp. 596–614.
25. Michon, J. A. *Dealing with Danger. Summary Report of a Workshop in the Traffic Research Centre*. The Netherlands, 1979.
26. Hart, Sandra, G. NASA-task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of Human Factors and Ergonomics Society Annual Meeting*, 2006, pp. 904–908.