Applied Ergonomics 58 (2017) 342–348

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Distraction and task engagement: How interesting and boring information impact driving performance and subjective and physiological responses

William J. Horrey [a,*], Mary F. Lesch [a], Angela Garabet [a], Lucinda Simmons [a], Rammohan Maikala [b]

[a] Liberty Mutual Research Institute for Safety, Hopkinton, MA, USA
[b] Providence Regional Medical Center, Providence Strategic and Management Services, Everett, WA, USA

ARTICLE INFO

Article history:
Received 20 February 2016
Received in revised form
1 June 2016
Accepted 21 July 2016
Available online 3 August 2016

Keywords:
Driver distraction
Task engagement/interest
Cerebral blood oxygenation
Heart rate
Pupillometry
Driving performance

ABSTRACT

As more devices and services are integrated into vehicles, drivers face new opportunities to perform additional tasks while driving. While many studies have explored the detrimental effects of varying task demands on driving performance, there has been little attention devoted to tasks that vary in terms of personal interest or investment—a quality we liken to the concept of task engagement. The purpose of this study was to explore the impact of task engagement on driving performance, subjective appraisals of performance and workload, and various physiological measurements. In this study, 31 participants (M = 37 yrs) completed three driving conditions in a driving simulator: listening to boring auditory material; listening to interesting material; and driving with no auditory material. Drivers were simultaneously monitored using near-infrared spectroscopy, heart monitoring and eye tracking systems. Drivers exhibited less variability in lane keeping and headway maintenance for both auditory conditions; however, response times to critical braking events were longer in the interesting audio condition. Drivers also perceived the interesting material to be less demanding and less complex, although the material was objectively matched for difficulty. Drivers showed a reduced concentration of cerebral oxygenated hemoglobin when listening to interesting material, compared to baseline and boring conditions, yet they exhibited superior recognition for this material. The practical implications, from a safety standpoint, are discussed.

1. Introduction

Today, drivers have access to more in-vehicle activities as new embedded and portable technologies become available and as more products and services are introduced that expand the connectivity of drivers and vehicles. While these devices and related activities afford drivers enhanced convenience and productivity, there are obvious safety concerns to the extent that they detract attention from the driving task, resulting in driver distraction (e.g., Regan et al., 2009).

Performance of a task, such as driving, is generally demanding of attentional resources and successful performance depends in part on the amount of resources demanded (e.g., difficulty) and the availability of resources to meet those needs (i.e., capacity; e.g., Wickens and Hollands, 2000). For very difficult tasks, there may be insufficient resources to accomplish the task and performance may suffer as a consequence. Likewise, when multiple tasks are performed concurrently, available resources will deplete more rapidly. Several studies in the driving context have shown that more demanding secondary tasks result in greater performance decrements in driving-related tasks (e.g., Briem and Hedman, 1995; Patten et al., 2004; Angell et al., 2006).

Tasks employed in experimental dual-task situations mobilize a similar set of underlying cognitive mechanisms as more naturalistic tasks (e.g., perception and information processing; use of working memory). However, they often fail to capture any semblance of personal investment or interest on the part of the participant. For example, performing mental arithmetic is an effective means of mobilizing cognitive resources; however, it does not constitute an activity that people would willfully engage in under normal

* Corresponding author.
E-mail address: william.horrey@libertymutual.com (W.J. Horrey).

http://dx.doi.org/10.1016/j.apergo.2016.07.011

circumstances. That is, motivation and interest in the to-be-performed tasks, irrespective of task difficulty, are likely to influence the individual's attentional allocation policy when performing concurrent activities. We refer to this as the quality of task *engagement*.

2. What is engagement?

The definition of engagement varies as a function of domain. Some refer to it as the overt or covert allocation of attention (Tops and Boksem, 2010), whereas others describe engagement by its propensity to attract and hold our attention (Chapman, 1997, 2003; Pintrich and De Groot, 1990) or even as a state in which an individual is so involved in an activity that all competing influences are blocked out (e.g., Csikszentmihalyi, 1990). While these approaches focus more on the outcome of engagement, O'Brien and Toms (2008) proposed a model of task engagement that focused on the properties of a task that would compel more or less engagement, including the degree to which tasks are challenging, interactive, rich in feedback, aesthetically pleasing, enduring, and varied or novel. With respect to tasks involving spoken auditory material, engagement can also be driven by many factors related to the gender, prosody and emotional tone of the speaker, among other factors (Nass and Brave, 2005). Interest in a particular task is assumed to be underscored by a positive affective response to the material and intrinsic motivation to perform the task. In the area of knowledge acquisition and skill development, level of interest is considered a central determinant of how we select and persist in processing certain types of information to the exclusion of others (e.g., Hidi, 1990). In the driving context, many of these factors likely influence users' willingness to perform potentially distracting tasks while driving (e.g., Lerner and Boyd, 2005).

2.1. Multitasking, engagement, and driving

Level of engagement can be influenced by a number of factors, regardless of task difficulty. As such, engagement may be an important factor in determining the driver's resource allocation policy over and above what might be required for successful task performance. Drivers may devote more attentional resources to activities that are considered engaging over ones that are less so and this could yield more detrimental implications for performance. Many studies of multitasking while driving have employed tasks that vary in terms of difficulty, but few have paid any attention to the individual's level of engagement or interest in the tasks.

For example, Dula et al. (2011) found that drivers in an emotional conversation engaged in more dangerous driving behaviors than drivers in a mundane conversation or a no-conversation condition. While these authors did not control for task difficulty (it was not a focus of their study), the study is intriguing because it underscores the potential for tasks that have certain attributes, such as those noted above, to impart different performance outcomes.

In earlier work, Horrey et al. (2009) examined two tasks that differed on many of the criteria for engagement (per O'Brien and Toms, 2008). The results suggested that a task that was considered to be more engaging (a twenty questions guessing game) led to worse performance outcomes than a less-engaging mental arithmetic task, although drivers rated their driving performance as superior with the engaging task. Although this dissociation was intriguing, the two tasks employed were quite different, in spite of attempts to model them after the criteria for task engagement, such that the observed pattern of results could have been influenced by differences in task structure, difficulty, and/or complexity (as seen in other experiments).

There have been no efforts to delineate the role and influence of task engagement on driving performance while controlling for task difficulty and task structure (e.g., modality, response demands). The current study purports to address this knowledge gap, with the specific aims of examining (1) performance implications of, and (2) physiological and subjective responses to, auditory material that was more interesting or engaging than mundane information of comparable difficulty.

2.2. Current study

Drivers in a driving simulator were exposed to boring (low engagement) auditory material, interesting (high engagement) material, or no auditory material (baseline). We examined driving performance along several dimensions of vehicle control and subjective responses. For the latter, we examined subjective ratings of workload and performance, self-reported interest in the material as well as post-trial recognition for presented material (as a potential index of the depth of processing; e.g., Craik and Lockhart, 1972). To the extent that more engaging material incurred more attention from drivers, we expected that performance would more adversely impacted and workload would be rated higher compared to boring material and baseline conditions.

In addition to performance implications, we were also interested in examining the physiological response to material of varying levels of engagement. We employed several physiological measurements. First, measures of cerebral hemoglobin oxygenation were monitored and assessed using an optical technique called near-infrared spectroscopy (NIRS). Previous research has also demonstrated that mental workload associated with visual, motor, and auditory stimuli evokes changes in regional oxidative metabolism of the brain (e.g., Roland, 1993; Derosière et al., 2013). The influence of mental workload on optical spectroscopy-derived responses during driving has increasingly being investigated (Harada et al., 2007; Li et al., 2009; Liu et al., 2012; Yoshino et al., 2013). For example, Harada et al. (2007) found that the concentration of oxygenated hemoglobin (O_2Hb) increased, with a concomitant decline in deoxygenated hemoglobin, when driving was compared to resting conditions (i.e., with increased workload). Similarly, Liu et al. (2012) demonstrated the sensitivity of NIRS-derived O_2Hb in the frontal lobe while driving under intrinsic and extrinsic cognitive load. However, none of the studies investigated the influence of listening to boring vs engaging material on cerebral responses while driving. To the extent that more engaging material could elicit the willful allocation of more cognitive resources (i.e., attention), we expected that the NIRS outcomes would reveal a greater mobilization of oxygenated hemoglobin than boring material, which would be more likely to be disregarded.

Additionally, we employed valid and reliable physiological measures of workload, including various pupil and heart rate parameters (e.g., O'Donnell and Eggemeier, 1986; Kramer, 1991; Sakamoto et al., 2009; Recarte et al., 2008; Beatty, 1982). For example, Mehler et al. (2012) found that heart rate increased with level of cognitive demand in drivers engaged in a concurrent working memory task. Similarly, Kun et al. (2013) found that pupil diameter was sensitive to variations in cognitive load associated with different aspects of dialogue.

3. Method

3.1. Participants

Thirty one drivers (aged 25 to 55, M = 37.0 yrs, SD = 8.5), 18 men and 13 women, participated in the study. They were screened for fluency in English, absence of self-reported hearing difficulties, visual function (visual acuity corrected to minimum 20/40), and low susceptibility to motion sickness. All participants were native English speakers. All drivers had a valid US Driver's License. They were paid $20 per hour for their participation.

3.2. Materials

Auditory Stimuli. Written material was gathered and modified from a number of news and online sources and initially classified by investigators as being either boring or interesting (see Table 1 for examples). 73 stimuli were then transcribed to audio clips of varying duration (range 12–38 s; M = 24 s) using an iTalk microphone (Griffin Technology, Nashville, TN) and iPod (Apple, Cupertino, CA).

Pilot Study and Stimuli Selection. In a pilot study, twenty-two participants (M = 39.3 yrs, SD = 10.6; balanced by gender), who did not participate in the main study, listened to the audio clips played through headphones. The clips were randomly arranged into two blocks, each lasting approximately 35 min. During each audio clip, participants made continuous ratings of their degree of interest in the material using a spring-dial controller. Turning the dial more to the right indicated increasing levels of interest and more to the left indicated increasing levels of boredom. Following each clip, participants made several additional ratings using 7-point Likert scales, including how likely they would be to attend to the material if they heard it, how entertaining they found the material, whether the clip had a conversational tone, and how difficult the material was to understand and how difficult it was to follow the language and structure. The audio clips were also scored using an objective measure, the Flesch-Kincaid Grade Level, which grades the difficulty of material according to the sentence length, word and syllable count.

The subjective ratings clearly corroborated the experimenters' pre-selection of boring and interesting material. The continuous measures of interest, scaled from −7 to +7, scored lower for Boring (M = −0.25) than Interesting items (M = 2.35; $t(21) = 8.5$, $p < 0.001$). Other ratings substantiated these differences; "if you heard this passage, how likely would you attend it" (Boring: M = 3.1; Interesting: M = 6.0, $t(21) = 15.7$, $p < 0.001$) and "how entertaining was the passage" (Boring: M = 2.6; Interesting: = 5.7, $t(21) = 17.7$, $p < 0.001$).

For the main study, we excluded audio clips that had inconsistent ratings; i.e., any boring items with scores that exceeded those of any interesting items and vice versa. The two groups of stimuli were then matched according to the objective measure of difficulty (grade-level; Boring: M = 12.8, SD = 2.4; Interesting: M = 12.5, SD = 3.1; $t(63) = 0.5$, $p = 0.63$). The final set included 65 audio clips, comprising 32 boring (range = 13–32s; M = 22.6s) and 33 interesting (range 11–34s, M = 21.9s) clips.

Driving Simulator. The fixed-base driving simulator (Realtime Technologies Inc., Royal Oak, MI) consisted of an open-cab vehicle mock up and three 46-inch widescreen LCD displays, subtending approximately 130 degrees of forward visual field. Data were collected at 60 Hz. The various driving environments and traffic scenarios were generated and coordinated using RTI SimCreator and SimVista software.

Optical Spectroscopy. A spatially resolved continuous wave near-infrared spectroscopy (NIRS) system (NIRO-300, Hamamatsu Photonics, Japan) was utilized to measure bilateral oxygenation from the prefrontal region. Two optical sensors were placed symmetrically on the anterior right and left frontal lobe 2–4 cm from the mid-line of the forehead and 2 cm above the supra-orbital ridge (Maikala et al., 2005). A dark colored tensor bandage was wrapped around the head to prevent ambient light entering each sensor. Measurements were recorded at 6 Hz and logged in relation to a datum created during system initialization, when participants were in a resting state.

Heart Rate. Heart rate was measured continuously at 200 Hz using a finger cuff system (Finometer, Finapres Medical Systems BV, The Netherlands) fitted to the middle or ring finger of the right hand. The heart rate monitor was calibrated at the start of each session.

Eye Tracker. A head mounted Eye Tracking System (ETS) (iViewX, SensoMotroric Instruments) was used to collect pupil diameter and eye movement data at a rate of 50 Hz. The system was mounted to a bicycle helmet to increase stability and to avoid interference with the placement of the NIRS sensors.

3.3. Procedure

Prior to the experimental session, participants were screened via online and phone survey for the minimum study requirements, including fluency in English and absence of self-reported hearing loss, and for susceptibility to simulator sickness. At the start of the 2.5-hour session, participants completed the informed consent, demographic and driving history questionnaire, motion sickness history and rating of symptoms related to simulator sickness. Visual function (acuity and color blindness) was assessed with a Titmus Vision Tester (Titmus Optical Inc., Chester, VA). Once seated in the driving simulator participants were fitted with the NIRS, heart rate

Table 1
Sample boring and interesting items.

Stimuli	Grade	Word count
Boring		
1) A local tree grower produces more than 50 species of trees and shrubs for planting on public and private land. The objective of the program is to provide low-cost plants from local sources to inspire landowners to enhance the state's environment for future generations. The tree nursery also offers a few non-native species.	10.4	54
2) From October 30 until December 3, grocers from New Hampshire to New Jersey participated in the Food for Friends campaign. Associates and customers donated more than $1.2 million through the sale of paper turkeys and fundraising events held at the stores. Together with corporate matches, the grocers will donate more than $1.4 million to local and regional food banks.	12.4	60
Interesting		
1) A local gym recently introduced "human barbells", hiring five men of various sizes (including two dwarfs) that customers could use for weights instead of the iron. One advantage of the human barbells is that, on request, they shout encouragement to the customer with each lift. The largest of the five is a 37-year-old, 340-pound man.	10.4	56
2) Four apparently quite bored people were arrested after a Chili's restaurant burglar alarm sounded at 4:30 a.m. According to police, the four intended to remove and steal the large chili on the restaurant's sign, using a hacksaw and power drill. To power the drill, they strung extension cords together running to the nearest outlet they could find, which was 470 feet away, across four lanes of highway and through a parking lot.	12.4	73

Note: these stimuli were modified from news stories gathered from dec.ny.gov/press, vineyardgazette.com, telegraph.co.uk, and newsoftheweird.com. Grade represents the Flesch-Kincaid Grade Level (described below in text).

monitor, and eye tracker and were introduced to the driving environments and tasks.

All drives consisted of 2-lane straight rural roads. At the beginning of each drive participants were instructed to accelerate until they approached a lead vehicle (travelling at a speed that varied randomly between 30 and 40 mph) and then maintain a close and constant following distance. At random intervals ranging from 30 to 60s, the lead vehicle abruptly decelerated and then resumed its normal speed. Participants were instructed to avoid hitting the lead vehicle by applying the brake and then to reestablish the close following distance.

In addition to the braking events, at random intervals (range 15–30s), the rear turn signal light of the lead vehicle illuminated for 2 s. Whenever this occurred, participants were instructed to activate their turn signal as fast as possible, in the opposite direction (e.g., if the lead vehicle signaled left, the driver engaged their right turn signal). This task was implemented to raise driver's overall workload and increase the challenge of the driving task.

Two short practice drives (approximately 4 min) were administered to allow participants to acclimatize to the control dynamics of the simulator and to practice each driving task without any additional auditory tasks. Drivers then completed 6 blocks of driving, each lasting approximately 8 min. These comprised 2 blocks each of: baseline (no audio recordings), boring audio clips and interesting audio clips. The two latter conditions used the stimuli selected from the pilot study. The order of the blocks was counterbalanced across participant and each block was balanced for total audio clip duration. Audio clips were presented on speakers mounted next to the simulator and played consecutively with a short interval (2 s) between each clip. Participants were deliberately not given any specific instructions concerning the auditory stimuli. After each driving block, drivers completed a modified NASA TLX (Hart and Staveland, 1988) which assessed drivers' ratings of mental demand, physical demand, time pressure, frustration, mental effort, as well as ratings of their performance on each of the vehicle following and turn signal tasks.

At the end of experiment, participants completed several questionnaires and post-drive symptoms of simulator sickness. Drivers were also given a recognition test for a subset of boring and interesting audio clips; they were asked to indicate whether or not they heard the clip and to indicate their confidence in their response. Five distractor passages were included (i.e., ones that were not presented during the driving blocks; mean grade level = 13.0). Additionally, drivers rated passages along several dimensions (likelihood of attending, entertainment value, conversational tone, comprehension, language and structure). A brief working memory test of digit-span was also administered, based on the Wechsler-Adult Intelligence Test. Finally, drivers were thanked, debriefed and remunerated for their participation.

4. Results

The design of the study included a one-way factor of driving condition with three levels: driving with no auditory stimuli (Baseline), driving while listening to boring auditory clips (Boring), and driving while listening to interesting auditory clips (Interesting). The various behavioral and physiological measures were analyzed using linear mixed models (LMM) in SPSS 20.0 according to the three levels of driving condition. In some cases, equipment malfunction and/or missing data for some of the data streams led to a smaller N for some comparisons and, consequently, some variation in the degrees of freedom. This predominantly impacted the physiological measurements.

4.1. Performance measures

Vehicle control. Several measures of vehicle control were recorded, including the average and standard deviation of headway distance as measured in meters from the bumper of the vehicle ahead to the simulator vehicle's bumper. The variability in lane position was the standard deviation of the lateral offset from road center (in meters). Similarly, the variability of speed (in meters per second) was taken as an index of speed control. All of these vehicle control measurements were aggregated at the block level.

The LMM revealed that car following distances were larger while listening to interesting material ($\underline{M} = 15.6$ m), compared to the boring ($\underline{M} = 14.9$ m) or baseline conditions ($\underline{M} = 14.6$ m; $\underline{F}(2,60) = 4.66$, p = 0.01, Cohen's $f^2 = 0.01$). Furthermore, the variability in following distances was reduced for both the boring ($\underline{M} = 3.77$ m) and interesting ($\underline{M} = 3.80$ m) auditory conditions compared to baseline ($\underline{M} = 5.90$ m; $\underline{F}(2,60) = 83.8$, p < 0.01, $f^2 = 0.40$). With respect to speed control, there was greater variability in vehicle speed in the baseline ($\underline{M} = 6.29$ mph) compared to the boring ($\underline{M} = 4.57$ mph) and interesting conditions ($\underline{M} = 4.44$ mph; $\underline{F}(2,60) = 376.8$, $\underline{p} < 0.001$, $f^2 = 1.48$). Lastly, drivers had significantly more variability in their lane keeping performance in the baseline condition ($\underline{M} = 0.16$ m) compared to the boring ($\underline{M} = 0.11$ m) and interesting ($\underline{M} = 0.10$ m) conditions ($\underline{F}(2,60) = 99.3$, $\underline{p} < 0.001$, $f^2 = 1.12$).

Braking and turn signal tasks. For each lead vehicle braking event, brake response time was measured from the start of the lead vehicle's deceleration until the first significant brake depression by the participant (defined as a 5% deflection). Prior to analysis, we removed unrealistically short response times (<200 m) that likely represented natural braking responses to slow the vehicle as opposed to a reaction to the lead vehicle. Only 5 samples (0.3% of the data) were removed as a result. The analysis revealed that braking response times were longer for the interesting condition ($\underline{M} = 0.94$ s), compared to boring ($\underline{M} = 0.92$ s) and baseline conditions ($\underline{M} = 0.90$ s; $\underline{F}(2,60) = 3.19$, $\underline{p} < 0.05$, $f^2 = 0.02$).

The turn signal task response time was measured from the onset of the lead vehicle's turn signal (left or right) until the driver's response was made. Accuracy was the percentage of correct responses on this task. Response times for the turn signal task did not vary across the different conditions (Baseline, $\underline{M} = 1.06$ s; Boring: $\underline{M} = 1.07$ s; Interesting, $\underline{M} = 1.07$ s; $\underline{F}(2,60) = 0.76$, $\underline{p} = 0.48$, $f^2 = 0.002$), nor did signal accuracy, which tended to be high ($\underline{M} = 99\%$ in all conditions; $\underline{F}(2,60) = 0.24$, $\underline{p} = 0.79$, $f^2 = 0.003$).

4.1.1. Physiological measurements

The various physiological measurements are summarized in Table 2, aggregated at the block level. NIRS-derived oxygenated (O_2Hb) and deoxygenated hemoglobin (HHb) was measured by micromoles per litre multiplied by the optical path length of the NIRS sensors. We also aggregated the O_2Hb and HHb densities from the left and right hemispheres. The analysis revealed significant differences across the conditions, with the baseline and boring conditions showing higher concentrations of O_2Hb than the interesting condition. There were no differences in HHb across the conditions; likewise, there were no significant effects with respect to the variability of the two measures.

Pupil diameter was measured horizontally in pixels using the eye camera image. The average pupil diameter did not vary across conditions, however, there was less variability (standard deviation) in pixel diameter in the Boring and Interesting conditions, compared to the Baseline driving condition. Heart rate was determined by beats per minute. Average heart rate and the variability of heart rate did not differentiate the three conditions.

Table 2
Measures and statistics for the physiological measurements.

	Condition			Stat[a]
	Baseline	Boring	Interesting	
Cerebral hemoglobin oxygenation[b] (in μM cm)				
Average O_2Hb	0.59	0.56	0.09	$F(2,57) = 4.5$, $\mathbf{p = 0.015}$, $f^2 = 0.02$
St. Dev. O_2Hb	0.96	1.02	1.02	$F(2,57) = 0.99$, $p = 0.38$, $f^2 = 0.007$
Average HHb	0.04	0.03	−0.02	$F(2,57) = 0.2$, $p = 0.83$, $f^2 = 0.001$
St. Dev. HHb	0.60	0.62	0.60	$F(2,57) = 0.26$, $p = 0.78$, $f^2 = 0.001$
Pupil diameter (PD) (in pixels)				
Average PD	45.4	45.2	44.5	$F(2,59) = 0.19$, $p = 0.82$, $f^2 = 0.001$
St. Dev. PD	3.7	2.8	2.6	$F(2,59) = 15.0$, $\mathbf{p < 0.001}$, $f^2 = 0.05$
Heart rate (HR) (in bpm)				
Average HR	75.4	75.6	75.0	$F(2,50) = 1.9$, $p = 0.15$, $f^2 = 0.001$
St. Dev. HR	7.8	7.2	8.2	$F(2,48) = 2.4$, $p = 0.10$, $f^2 = 0.007$

Note: O_2Hb = oxygenated hemoglobin; HHb = deoxygenated hemoglobin; μM cm = (micromoles per litre × optical path length). bpm = beats per minute. Bold entries indicate statistical significance at $p < 0.05$.
[a] From LMM analysis.
[b] Values represent changes from system calibration and initialization, taken during resting state.

5. Recognition performance & subjective ratings

Recognition performance. Drivers' recognition of the auditory material showed some pronounced differences across the conditions ($F(2,63) = 16.1$, $p < 0.001$, $f^2 = 0.35$). For interesting material, they correctly identified 92.6% of the stimuli presented in the recall test. In comparison, they were only 63.3% accurate for boring material. Participants correctly rejected 84.7% of the catch trials (i.e., stimuli that had not been presented in the study). With respect to the catch trials, which included both interesting and boring material, participants were better able to reject unheard material for interesting (93.3% accuracy) than boring material (78.8%; $F(1,29) = 11.8$, $p = 0.002$, Cohen's $d = 0.57$).

Subjective workload ratings. Analysis of the unweighted composite workload score, based on the average of all of the sub-scales of the modified NASA TLX, revealed small but significant differences across the conditions ($F(2,56) = 6.5$, $p = 0.003$, $f^2 = 0.01$), with the Boring condition ($M = 45.3$) scoring higher workload ratings than the Baseline ($M = 43.3$) and Interesting condition ($M = 42.1$).

Given our interest in perceived performance, we examined subjective ratings of driving (car following) performance across the conditions using the performance subscale of the NASA TLX. The analysis did not show any differences between the conditions ($F(2,56) = 0.59$, $p = 0.56$, $f^2 = 0.003$; Baseline = 68.4; Boring = 67.2; Interesting = 66.1).

Subjective ratings of task engagement/difficulty. Lastly, we note that the post-experimental ratings of a subset of the stimuli used in the study verified our manipulation of interest and also corroborated the outcomes from the pilot study. Participants indicated, on a 7-point scale, that they would be more likely to attend to the interesting material ($F(1,29) = 98.3$, $p < 0.001$, $d = 0.88$; Boring = 3.5; Interesting = 5.3) and that the material was more entertaining ($F(1,29) = 120.8$, $p < 0.001$, $d = 2.1$; Boring = 2.7; Interesting = 5.2). Although the material was matched using objective measures of difficulty, participants rated the interesting material as being less difficult to understand ($F(1,29) = 12.4$, $p = 0.001$, $d = 0.39$; Boring = 2.0; Interesting = 1.7), although both conditions were rated as easy to understand.

6. Discussion

The current study aimed to examine the performance implications of, and the physiological and subjective responses to, auditory material that was more interesting or engaging than mundane information of comparable difficulty. Previous work in driving research has demonstrated the detrimental effects of difficult secondary activities (e.g., Briem and Hedman, 1995; Patten et al., 2004). While the level of demand or task difficulty can influence how challenging a task may be (and hence, how well it can be performed), it does not necessarily dictate how an individual will apportion their effort towards the task (Wickens and Hollands, 2000). It is also possible that the mobilization of attention might be easier with material that is more intrinsically interesting, since the material can grab attention (e.g., Chapman, 1997). Alternatively, participants might be motivated to attend to material that isn't intrinsically interesting (i.e., for work). Certainly, other intrinsic properties will influence how engaging a task might be and how an individual will apportion their attentional resources (e.g., O'Brien and Toms, 2008; Nass and Brave, 2005).

Some of the results from the current study corroborated previous findings concerning the performance of concurrent activities while driving. For example, under both auditory conditions, drivers exhibited less variability in lane keeping performance (e.g., Brookhuis et al., 1991) as well as speed control and headway maintenance. However, the turn signal task, though likely adding to the overall workload of the driving block, did not yield any differences across conditions, with performance being consistently high. Importantly for our purposes, the interesting auditory condition—here a proxy for increased task engagement—did lead to some different patterns of outcomes. For example, the delayed response times to the critical braking events were more pronounced in the interesting audio condition, compared to the baseline and boring conditions.

The physiological and subjective measures showed some interesting patterns. Generally, studies of have shown that more difficult cognitive tasks are associated with increased oxygenated hemoglobin and decreased deoxygenated hemoglobin (e.g., Derosière et al., 2013; León-Carrion et al., 2008; Villringer et al., 1993; Hoshi and Tamura, 1993). However, drivers in our study showed reduced cerebral oxygenated hemoglobin when listening to interesting material, compared to baseline and boring conditions, suggesting that engaging tasks do not necessarily command a greater mobilization of cognitive resources. Yet, driver's superior recognition of the interesting material seemed to suggest a greater depth of processing of the material (e.g., Craik and Lockhart, 1972) or that interesting material has some intrinsic properties that facilitate its processing (e.g., promotes easier connections to long-term memory). This pattern suggests that, from a hemodynamic perspective, drivers are able to do “more with less” for engaging, interesting material. In a similar vein, Shirey and Reynolds (1988) found that, compared to boring material, interesting material was learned better albeit with less attention allocated (measured by the time spent reading the text). We do grant, however, that the catch trial data in the current study suggested some advantages of the interesting material with respect to discriminability (i.e., it was easier to distinguish between unheard interesting items than boring). More work is needed to further parse these effects. While, collectively, these results could indicate some decided advantages for information processing and the preservation of attentional resources, we underscore the adverse performance implications on the critical braking task in the interesting condition—an outcome that stresses some of the concerns over such activities while driving.

The influence of task engagement on physiological responses was further substantiated with the observed variability in pupil diameter, where more variability was observed in the baseline condition compared to the two auditory task conditions (cf. Sakamoto et al., 2009). It is possible that this could be due to increased variability in scene luminance brought about by more scanning in the baseline condition—a possibility that can be explored further. Furthermore, heart rate responses were similar among the three conditions, in spite of variations in mental workload and differences in heart rate variability failed to reach conventional levels of statistical significance. Although many studies have documented different outcomes concerning heart rate and heart rate variability, Lee and Park (1990) demonstrated, in the same study, a reduced heart rate variability in light of increased mental load, but no influence on heart rate.

Also noteworthy is the subjective responses towards the different material. Not surprisingly, participants rated the interesting material as more entertaining and attention-grabbing than the boring material. However, they also perceived the material to be easier and less complex, even though the material was matched for difficulty. It is possible that degree of interest in the stimuli could influence subjective appraisals of difficulty or that participants are sensitive to some of the apparent efficiencies gained in the processing of interesting material. Along these lines, the ratings of workload were lower for this condition than the other two. The reduced subjective workload ratings in the interesting condition are further corroborated with lower increase in cerebral oxygenation during listening to interesting material as compared to that of driving while listening to boring material. These cerebral findings are in contrast to Eastwood et al. (2012)'s reported relationship between boredom and attention, implying that although the current material was boring, participants in our study sustained their attention with greater effort to accomplish the driving task, thus experiencing increased mental workload, leading to greater increase in cerebral oxygenation.

6.1. Limitations

While the results are suggestive, the current study is not without limitations. We endeavored to match our interesting and boring material in terms of difficulty. To do so, we used an objective measure of grade (difficulty) level; however, this was based on a measure of reading comprehension. Although this takes into account variables that would also be expected to influence difficulty in auditory comprehension, it could be less sensitive to some of the nuances and complexities of language. Matching information according to difficulty, length, and a number of other factors (including tone and prosody) is very challenging and it stands that additional controlled work in this area is worthwhile. In addition to difficulty, we also tried to maintain consistency across boring and interesting audio clip in terms of prosody, intonation and other vocal features, not to mention the gender of the speaker. As noted by Nass and Brave (2005), these types of features can also influence engagement and reaction to spoken material and therefore represent areas of further study in this space.

With respect to task instructions, we did not offer any specific guidance regarding the auditory material. Given that the engaging properties of the interesting material could draw attention, it was thought that offering instructions might have altered the manner in which the information—in both conditions—was treated. However, in the absence of any instructions, drivers might have made their own interpretations regarding the auditory material and acted accordingly (e.g., predicting that a memory test would occur)—a strategy that could have superseded or watered-down the natural impact of interest/engagement.

Lastly, we did not synchronize the presentation of auditory material to critical events, instead allowing the events to occur according to random scheduling. It is possible that the observed effects would become even more pronounced when engaging material coincides with critical events. Indeed, in many cases the observed effects were small, albeit statistically significant. Along these lines, we also note that the experience of interest and boredom is not a discrete or singular event; rather it varies continuously, rising and falling with the momentary information (e.g., consider a narrative that slowly builds up to something unexpected) as well as according to other individual and situational factors. Understanding these temporal variations and their impact on attention and performance is certainly an area for further investigation.

6.2. Practical implications & conclusions

In real world driving situations, whether being carried out for personal or business reasons, drivers are afforded many opportunities to interact with external devices and information. However, unlike in a laboratory setting, drivers will be likely engaged in activities in which they are personally vested. This encompasses activities that might be of personal interest as well as those that might be work-related. It is important that researchers and safety practitioners understand how these different types of tasks or information can impact behavior, decision-making and performance. In the current effort, we have tried to differentiate some of the performance implications of task interest and engagement while controlling for difficulty, showing that interesting material present—*even without any specific instructions concerning it*—can draw drivers' attention and lead to delayed braking responses to critical road hazards. However, there needs be more work done to understand the various motivational factors that can impact the decision-making process concerning non-driving related activities (e.g., Lerner and Boyd, 2005). These issues will be of continued importance as new and advanced autonomous systems become more prevalent in vehicles (e.g., Merat et al., 2014; Gold et al., 2013). With such systems, drivers will be in full control of the vehicle at times and will be monitors of these systems at other times; it is important to understand how non-driving activities can impact drivers' readiness to intervene or resume control of the vehicle—especially for tasks that are more engaging.

Furthermore, the results suggested that certain measures of cerebral hemoglobin oxygenation might be sensitive to variations in observers' interest or engagement to auditory material, even though other physiological measures such as pupil diameter and heart rate were not sensitive. Certainly more work in this area is merited in order to corroborate and expand the current findings; however, it does highlight the potential for new ways of physiologic monitoring in situ. Compared to other imaging approaches (e.g., fMRI), NIRS can be more portable, modestly prices and is easy to implement, offering some appealing features for use in research as well as operational settings. That said, we cede that many more technological advances are necessary before this technique would be practical outside of a controlled laboratory. For example, equipment that did not require careful set up and calibration—possibly contact-free—would have much more appeal in driver state sensing applications in private vehicles. As a final note, the current results shed some new and potentially important insights into the role and influence of task engagement on drivers' responses and performance and we hope that the current outcomes will foster new interest and effort in this area.

Acknowledgements

A subset of data was presented at the 19th Triennial Congress of the International Ergonomics Association, held in Melbourne, Australia, 2015. We are grateful to Chris Wickens, Niall O'Brien, Jacob Banks, Amanda Rivard, Susan Jeffries, Sarah Hong, Marvin Dainoff, Lawrence Hettinger, and Maurizio Trippolini for their help and guidance on this project or manuscript. Three anonymous reviewers also provided valuable feedback.

References

Angell, L., Auflick, J., Austria, P.A., Kochhar, D., Tijerina, L., Biever, W., et al., 2006. Driver Workload Metrics Task 2 Final Report and Appendices. National Highway Traffic Safety Administration, Washington, DC (Report No. DOT HS 810 635).
Beatty, J., 1982. Task-evoked pupillary responses, processing load, and the structure of processing resources. Psychol. Bull. 91 (2), 276–292.
Briem, V., Hedman, L.R., 1995. Behavioral effects of mobile telephone use during simulated driving. Ergonomics 38 (12), 2536–2562.
Brookhuis, K.A., de Vries, G., de Waard, D., 1991. The effects of mobile telephoning on driving performance. Accid. Analysis Prev. 23, 309–316.
Chapman, P., 1997. Models of Engagement: Intrinsically Motivated Interaction with Multimedia Learning Software. Unpublished Masters Thesis. University of Waterloo, Waterloo, Canada.
Chapman, E., 2003. Alternative approaches to assessing student engagement rates. Pract. Assess. Res. Eval. 8 (13). Retrieved May 22, 2008 from. http://PAREonline.net/getvn.asp?v=8&n=13.
Recarte, M.Á., Pérez, E., Conchillo, Á., Nunes, L.M., 2008. Mental workload and visual impairment: differences between pupil, blink, and subjective rating. Span. J. Psychol. 11 (02), 374–385.
Craik, F.I., Lockhart, R.S., 1972. Levels of processing: a framework for memory research. J. Verbal Learn. Verbal Behav. 11 (6), 671–684.
Csikszentmihalyi, M., 1990. Flow: the Psychology of Optimal Experience. Harper and Sons, New York.
Derosière, G., Mandrick, K., Dray, G., Ward, T.E., Perrey, S., 2013. NIRS-measured prefrontal cortex activity in neuroergonomics: strengths and weaknesses. Front. Hum. Neurosci. 7, 583.
Dula, C.S., Martin, B.A., Fox, R.T., Leonard, R.L., 2011. Differing types of cellular phone conversations and dangerous driving. Accid. Analysis Prev. 43 (1), 187–193.
Eastwood, J.D., Frischen, A., Fenske, M.J., Smilek, D., 2012. The unengaged mind defining boredom in terms of attention. Perspect. Psychol. Sci. 7 (5), 482–495.
Gold, C., Damböck, D., Lorenz, L., Bengler, K., 2013. "Take over!" How long does it take to get the driver back into the loop?. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 57, No. 1. SAGE Publications, pp. 1938–1942.
Harada, H., Nashihara, H., Morozumi, K., Ota, H., Hatakeyama, E., 2007. A comparison of cerebral activity in the prefrontal region between young adults and the elderly while driving. J. Physiol. Anthropol. 26 (3), 409–414.
Hart, S.G., Staveland, L.E., 1988. Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. North-Holland, Amsterdam, pp. 139–183.
Hidi, S., 1990. Interest and its contribution as a mental resource for learning. Rev. Educ. Res. 60, 549–571.
Horrey, W.J., Lesch, M.F., Garabet, A., 2009. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. J. Saf. Res. 40 (1), 7–12.
Hoshi, Y., Tamura, M., 1993. Detection of dynamic changes in cerebral oxygenation coupled to neuronal function during mental work in man. Neurosci. Lett. 150, 5–8.
Kramer, A.F., 1991. Physiological metrics of mental workload: a review of recent progress. In: Damos, D.L. (Ed.), Multiple-task Performance. Taylor & Francis, London, pp. 279–328.
Kun, A.L., Palinko, O., Medenica, Z., Heeman, P.A., 2013. On the feasibility of using pupil diameter to estimate cognitive load changes for in-vehicle spoken dialogues. In: INTERSPEECH, pp. 3766–3770.
Lee, D.H., Park, K.S., 1990. Multivariate analysis of mental and physical load components in sinus arrhythmia scores. Ergonomics 33 (1), 35–47.
León-Carrion, J., Damas-López, J., Martín-Rodríguez, J.F., Domínguez-Roldán, J.M., Murillo-Cabezas, F., Barroso, Y., Martin, J.M., Domínguez-Morales, M.R., 2008. The hemodynamics of cognitive control: the level of concentration of oxygenated hemoglobin in the superior prefrontal cortex varies as a function of performance in a modified Stroop task. Behav. Brain Res. 193 (2), 248–256.
Lerner, N., Boyd, S., 2005. Task Report: On-road Study of Willingness to Engage in Distracting Tasks (Report No. DOT HS 809 863). National Highway Traffic Safety Administration, Washington, DC.
Li, Z., Zhang, M., Zhang, X., Dai, S., Yu, X., Wang, Y., 2009. Assessment of cerebral oxygenation during prolonged simulated driving using near infrared spectroscopy: its implications for fatigue development. Eur. J. Appl. Physiol. 107, 281–287.
Liu, T., Saito, H., Oi, M., 2012. Distinctive activation patterns under intrinsically versus extrinsically driven cognitive loads in prefrontal cortex: a near-infrared spectroscopy study using a driving video game. Neurosci. Lett. 506, 220–224.
Maikala, R.V., King, S., Bhambhani, Y.N., 2005. Cerebral oxygenation and blood volume responses to seated whole-body vibration. Eur. J. Appl. Physiol. 95, 447–453.
Mehler, B., Reimer, B., Coughlin, J.F., 2012. Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task an on-road study across three age groups. Hum. Factors J. Hum. Factors Ergon. Soc. 54 (3), 396–412.
Merat, N., Jamson, A.H., Lai, F.C., Daly, M., Carsten, O.M., 2014. Transition to manual: driver behaviour when resuming control from a highly automated vehicle. Transp. Res. Part F Traffic Psychol. Behav. 27, 274–282.
Nass, C.I., Brave, S., 2005. Wired for Speech: How Voice Activates and Advances the Human-computer Relationship. MIT press, Cambridge.
O'Donnell, R., Eggemeier, F.T., 1986. Workload assessment methodology. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Human Performance. Wiley-Interscience.
O'Brien, H.L., Toms, E.G., 2008. What is user engagement? A conceptual framework for defining user engagement with technology. J. Am. Soc. Inf. Sci. Technol. 59 (6), 938–955.
Patten, C.J., Kircher, A., Ostlund, J., Nilsson, L., 2004. Using mobile telephones: cognitive workload and attention resource allocation. Accid. Analysis Prev. 36 (3), 341–350.
Pintrich, P.R., De Groot, E., 1990. Motivational and self-regulated learning components of classroom academic performance. J. Educ. Psychol. 82 (1), 33–50.
Regan, M.A., Lee, J.D., Young, K.L., 2009. Driver Distraction: Theory, Effects, and Mitigation. CRC Press, Boca Raton, FL.
Roland, P.E., 1993. Brain Activation. Wiley-Liss, New York.
Sakamoto, K., Aoyama, S., Asahara, S., Mizushina, H., Kaneko, H., 2009. Relationship between Emotional State and Pupil Diameter Variability under Various Types of Workload Stress. In Ergonomics and Health Aspects of Work with Computers. Springer Berlin Heidelberg, pp. 177–185.
Shirey, L.L., Reynolds, R.E., 1988. Effect of interest on attention and learning. J. Educ. Psychol. 80 (2), 159–166.
Tops, M., Boksem, M.A.S., 2010. Absorbed in the task: personality measures predict engagement during task performance as tracked by error negativity and asymmetrical frontal activity. Cognitive, Affect. Behav. Neurosci. 10 (4), 441–453.
Villringer, A., Planck, J., Hock, C., Schleinkofer, L., Dirnagl, U., 1993. Near infrared spectroscopy (NIRS): a new tool to study hemodynamic changes during activation of brain function in human adults. Neurosci. Lett. 154, 101–104.
Wickens, C.D., Hollands, J.G., 2000. Engineering Psychology and Human Performance, third ed. Prentice Hall, Upper Saddle River, NJ.
Yoshino, K., Oka, N., Yamamoto, K., Takahashi, H., Kato, T., 2013. Functional brain imaging using near infrared spectroscopy during actual driving on an expressway. Front. Hum. Neurosci. 7 (882), 1–16.