

Manipulating Drivers' Mental Workload: Neuroergonomic Evaluation of the Speed Regulation N-Back Task Using NASA-TLX and Auditory P3a

Nikol Figalová*

nikol.figalova@uni-ulm.de
Clinical and Health Psychology, Inst.
of Psychology and Education, Ulm
University
Ulm, Germany

Olga Pollatos

Clinical and Health Psychology, Inst.
of Psychology and Education, Ulm
University
Ulm, Germany

Jürgen Pichen*

juergen.pichen@uni-ulm.de
Human Factors, Inst. of Psychology
and Education, Ulm University
Ulm, Germany

Lewis Chuang

Human and Technology, Inst. for
Media Research, Faculty of
Humanities, Chemnitz University of
Technology
Chemnitz, Germany

Vanchha Chandrayan

Human Factors, Inst. of Psychology
and Education, Ulm University
Ulm, Germany

Martin Baumann

Human Factors, Inst. of Psychology
and Education, Ulm University
Ulm, Germany



Figure 1: The experimental setup: A participant is conducting the modified speed regulation n-back task in a driving simulator while his electroencephalography (EEG) is being recorded.

ABSTRACT

Manipulating MW in driving simulator studies without the need to introduce a non-driving-related task remains challenging. This study aims to empirically evaluate the modified speed regulation n-back task, a tool to manipulate drivers' MW. Our experiment involved 23 participants who experienced a 0-back and 2-back driving condition, with task-irrelevant novel environmental sounds used to elicit P3a event-related potentials. Results indicate that the 2-back condition was perceived as more demanding, evidenced by higher NASA-TLX scores (overall score, mental and temporal demand, effort, frustration). The mean P3a amplitude was diminished during

the 2-back condition compared to the 0-back condition, suggesting that drivers experienced higher MW and had fewer resources available to process the novel environmental sounds. This study provides empirical evidence indicating that the speed regulation n-back task could be a valid, effective, and reproducible method to manipulate MW in driving research.

KEYWORDS

mental workload, driving, speed regulation n-back task, auditory P3a, neuroergonomics

ACM Reference Format:

Nikol Figalová, Jürgen Pichen, Vanchha Chandrayan, Olga Pollatos, Lewis Chuang, and Martin Baumann. 2023. Manipulating Drivers' Mental Workload: Neuroergonomic Evaluation of the Speed Regulation N-Back Task Using NASA-TLX and Auditory P3a. In *15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '23 Adjunct)*, September 18–22, 2023, Ingolstadt, Germany. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3581961.3609887>

1 INTRODUCTION

There are various methods to manipulate mental workload (MW) in the field of driving research. Traditionally, researchers would ask

*Both authors contributed equally to this research.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

AutomotiveUI '23 Adjunct, September 18–22, 2023, Ingolstadt, Germany

© 2023 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-0112-2/23/09.

<https://doi.org/10.1145/3581961.3609887>

participants to perform a non-driving related task (NDRT) while driving or monitoring an automated vehicle to manipulate the level of experienced MW. This approach can be described as a dual-task approach. Contrary, a single-task approach requires drivers to be involved only in one task, typically driving. In order to increase or decrease the demands of driving, researchers would manipulate environmental factors, such as traffic complexity [17].

Implementing MW manipulation within a single-task approach presents significant challenges. Researchers often modify the difficulty of the task based on their best understanding (for instance, by increasing traffic in the simulated scenario or reducing visibility) and presume the manipulation yields the intended effect. However, these manipulations are not standardized, making it problematic to evaluate or replicate the effects. Conversely, applying a standardized secondary NDRT to manipulate MW addresses issues of reproducibility and subjectivity in experimental designs. Yet, accurately estimating the distribution of participants' mental resources between primary and secondary tasks can be problematic. Moreover, it potentially undermines the realism of the driving experience [10].

Acquiring controlled, standardized MW manipulation within a single-task driving paradigm remains challenging. This paper aims to provide further empirical evidence for the impact of a single-task MW manipulation using a modified speed regulation n-back driving task proposed by Unni et al. [18]. In the present study, participants encountered a 0-back and 2-back condition within a driving simulator. We analysed self-reported MW (using NASA-TLX [8]) and objective MW (measured via P3a amplitude induced by task-irrelevant environmental sounds [14, 15]). To our best knowledge, this study is the first attempt to assess the speed regulation n-back task employing P3a event-related potential (ERP). Moreover, no self-reported assessment of perceived MW induced by the n-back task has been previously published.

Understanding the effect of MW manipulation via the speed regulation n-back task on the perceived cognitive load and P3a amplitude will provide valuable insights into the cognitive mechanisms underlying this task. Providing empirical evidence combining objective and subjective measures of MW will help to establish the speed-regulation n-back task as an objective, replicable way of MW manipulation in driving simulator studies.

1.1 Related Work

Mental workload (MW) refers to the amount of cognitive resources required to perform a specific task [4]. It's a measure of the mental effort expended, taking into account factors such as task complexity, information processing demands, and the cognitive capabilities of the individual. MW is measurable through various techniques, including self-reporting, performance-based assessments, and physiological measurements [3]. Using event-related potentials (ERPs) evoked by novel environmental auditory cues has been proposed to be an objective physiological measure of MW [5, 12, 14]. The authors of these studies jointly report that higher MW correlates with a lower amplitude of the P3a ERP component and vice versa. This effect is based on the multiple resources theory [21], which suggests that the environmental stimuli compete for limited processing

resources. When a high mental workload is induced, fewer attentional resources are left for processing the environmental sounds [2]. Higher MW, therefore, leads to decreased P3a amplitude. There is a discussion about whether other factors, such as arousal or top-down attention control, might be a factor influencing the P3a amplitude [1, 7]; nevertheless, robust empirical evidence supports the inverted relationship between MW and P3a amplitude due to the competition for resources.

Several authors studied the P3a amplitude in the driving research. Wester et al. compared the P3a amplitudes between stationary and driving conditions. Scheer et al. studied the effect of steering demands on P3a amplitude. Van der Heiden et al. compared the P3a amplitude in a stationary condition to manual and automated driving. Figalová et al. studied the effect of different levels of automation on the P3a amplitude. However, none of these studies used a standardised single-task method to manipulate MW.

An n-back task is a continuous performance task designed to measure a part of working memory and working memory capacity. The load factor n can be adjusted to make the task more or less difficult and, therefore, to manipulate MW. In the modified speed regulation n-back driving task proposed by Unni et al. [18], participants manually drive while presented with a series of speed signs. The task requires participants to continuously update, memorise, and recall previous speed signs while adjusting their current speed to match the sign displayed n -steps before. This paradigm has been used in other studies [9, 16, 18] and appears to be an efficient way to manipulate MW using a single-task approach.

1.2 Hypotheses

Participants experienced a low MW condition (0-back) and a high MW condition (2-back). Based on the inverted relationship between MW and P3a amplitude discussed in section 1.1, we formulated the following hypotheses:

- **H1:** The overall score of NASA-TLX is lower during the 0-back than during the 2-back condition.
- **H2:** The mean P3a amplitude is higher during the 0-back than during the 2-back condition.

2 METHODS

2.1 Participants

Our sample consisted of 23 participants (11 males and 12 females) with an average age of 23.83 years ($SD = 3.61$). All participants had normal or corrected-to-normal vision, no known neurological or psychiatric disease, and a valid German driving license (on average for $M = 3.77$ years, $SD = 3.48$). Participants provided informed consent and received 50 Euros as compensation to participate.

2.2 Apparatus and Procedure

The experiment took place in a highly immersive, fixed-base driving simulator with a realistic car mock-up (see Figure 1). The EEG was recorded using 32 active channels (ActiCAP by Brain Products) placed according to the international 10-20 system. We kept the impedance below $25k\Omega$. We also recorded electrodermal activity and electrocardiogram; however, the data are not presented here. The perceived cognitive workload was assessed using the NASA

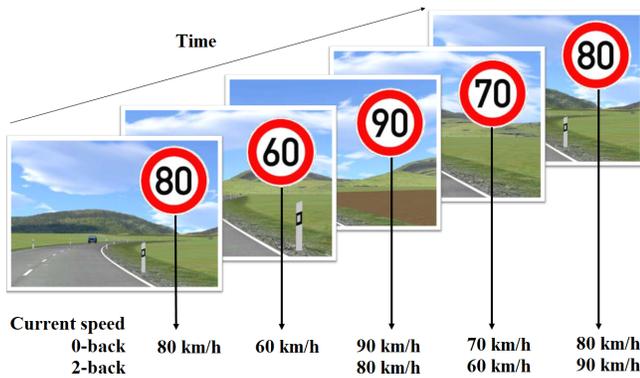


Figure 2: Example of the speed regulation n-back task.

Task Load Index scale (NASA-TLX; [8]). Each of the six dimensions was evaluated on a 100-point Likert scale. The total score was obtained as the average of the six dimensions.

First, participants learned how to operate the driving simulator and conduct the speed regulation n-back task. Afterwards, we installed the EEG and started the experiment. Participants filled in the NASA-TLX at the end of each condition. The order of the conditions was counterbalanced between participants.

2.3 Experimental task

Participants experienced speed regulation 0-back and 2-back conditions designed to manipulate MW using a within-subject design. Participants drove on a rural road, and a speed sign (from 60 to 100 km/h) was presented in a random order every 20 seconds on a head-up-like screen in front of them (see Figure 1). The position and size of the sign corresponded to a normal German speed sign. The sign was shown for 3 seconds. For the 0-back task, participants were instructed to adjust and maintain the speed according to the currently presented sign. For the 2-back task, participants had to adjust and maintain the speed that occurred two signs before and remember the sequence of the two following speed signs. Each condition lasted for approximately 17 minutes, and 52 signs were shown in each condition. The task is visualised in Figure 2.

2.4 Stimuli and Questionnaires

Participants were exposed to a passive oddball task, which consisted of three types of task-irrelevant auditory stimuli:

- (1) **Frequent distractors**, 450 presentations in each condition (probability of presentation: 71.43%);
- (2) **infrequent distractors**, 90 presentations in each condition (probability of presentation: 14.29%);
- (3) **novel environmental distractors**, 90 presentations in each condition (probability of presentation: 14.29%).

The frequent and infrequent distractors consisted of two tones (pure 700 Hz and 300 Hz tones); their presentation probability was counterbalanced across participants. The novel environmental distractors consisted of 30 complex sounds (e.g., human laughter, helicopter) [6]. Each stimulus was presented for 336 ms. The inter-stimulus interval was randomized and ranged from 1300 to 1700 ms.

2.5 EEG Pre-processing and Analysis

Data preprocessing followed the BeMoBil pipeline [11]. The data were submitted for adaptive mixture independent component analysis (AMICA). We classified the components using IClab [13] and retained the components that most likely originated from brain activity. Next, we filtered the data between 0.1 and 30 Hz and extracted epochs from -200 to 800 ms relative to the stimulus onset. Artifactual epochs were rejected. The preprocessed data were averaged for each channel and experimental condition. A difference wave was computed as the difference between the ERP elicited by the environmental distractor and the ERP elicited by the frequent distractor. Statistical analysis was conducted focusing on the Fz electrode using JASP 0.16.4, with the general level of significance set to 0.05.

3 RESULTS

3.1 Self-reported Mental Workload

Using a paired samples t-test, we found a difference in the **total NASA-TLX score** between 0-back ($M = 42.464$, $SD = 15.187$) and 2-back ($M = 58.478$, $SD = 14.466$) experimental conditions; $t(22) = 7.326$, $p < .001$, $d = 1.528$. Comparing the six dimensions of NASA-TLX, we found a difference in **mental demand** ($M = 36.957$, $SD = 21.253$ for 0-back; $M = 77.826$, $SD = 19.588$ for 2-back; $t(22) = 9.700$, $p < .001$, $d = 2.023$), **temporal demand** ($M = 39.348$, $SD = 21.443$ for 0-back; $M = 53.478$, $SD = 25.603$ for 2-back; $t(22) = 3.472$, $p = .002$, $d = 0.724$), **effort** ($M = 45.652$, $SD = 24.416$ for 0-back; $M = 73.261$, $SD = 20.204$ for 2-back; $t(22) = 6.260$, $p < .001$, $d = 1.305$), and **frustration** ($M = 29.565$, $SD = 31.315$ for 0-back; $M = 47.391$, $SD = 27.214$ for 2-back; $t(22) = 3.790$, $p < .001$, $d = 0.790$). The results are visualised in 3a.

The results suggest that the overall perceived mental workload was lower in the 0-back task than in the 2-back task. Moreover, this finding is consistent with the differences observed in the single factors of NASA-TLX. Therefore, we **accept H1**.

3.2 Mean P3a amplitude

We used the collapsed localizer method to quantify ERP component amplitudes. A grand average peak (GA) was determined at 328 ms post-stimulus. The mean amplitude was calculated between 303 and 353 ms post-stimulus (GA peak \pm 25 ms). Figure 3b presents the difference waves (environmental-frequent). We found a difference in the mean amplitude between 0-back ($M = 1.429$ μ V, $SD = 1.272$) and 2-back ($M = 2.471$ μ V, $SD = 2.471$) experimental conditions; $t(22) = 1.937$, $p = .033$, $d = 0.404$.

The results suggest that the P3a amplitude elicited by the novel environmental sounds was decreased with increased mental workload. Therefore, we **accept H2**.

4 DISCUSSION

This study investigated the effect of MW manipulation using the modified speed regulation n-back task proposed by Unni et al. [18]. We conducted a driving simulator experiment with 23 participants who experienced 0-back and 2-back conditions. We analysed the self-reported MW (using NASA-TLX [8]). Moreover, we studied the mean amplitude of P3a ERP components evoked by task-irrelevant

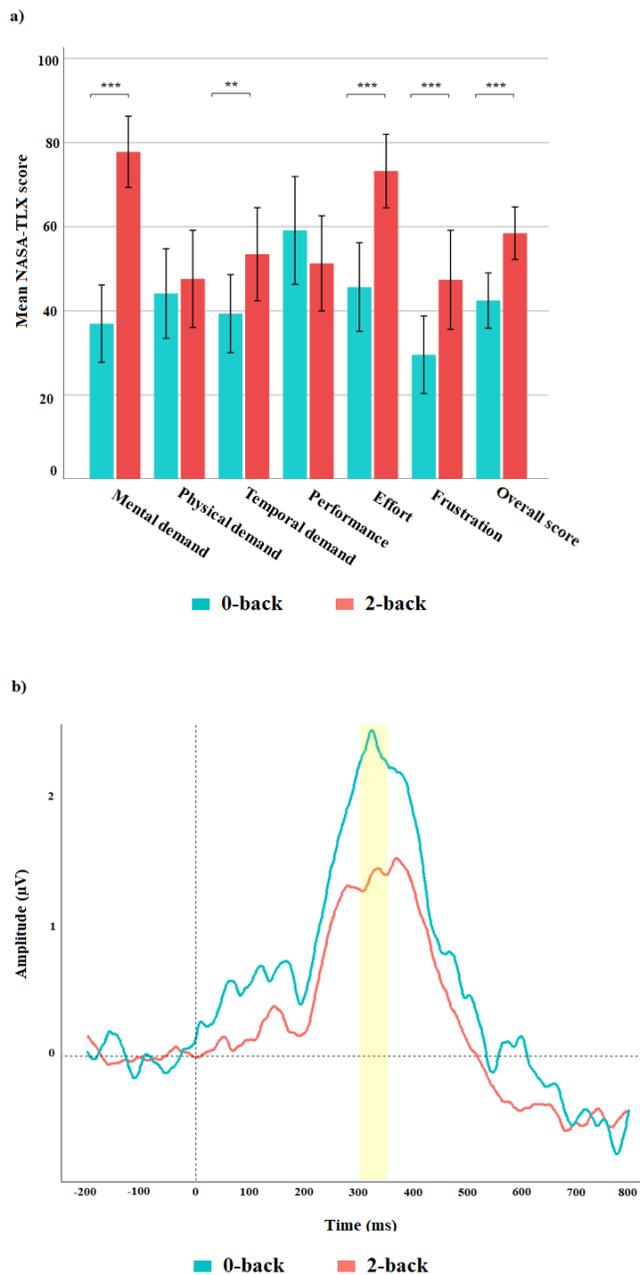


Figure 3: a) The mean score of the six dimensions of NASA-TLX. The error bars represent a 95% confidence interval. Significant differences are highlighted ($*** p \leq .001$, $** p \leq .01$). b) Difference wave (environmental novel distractor-frequent distractor) ERPs for the 0-back and 2-back experimental conditions at the Fz electrode. Mean amplitudes in the yellow-highlighted area (303 to 353 ms) were submitted for statistical comparisons.

novel sounds, which served as a proxy measure of objective MW [12, 15].

The subjective assessment of MW using the overall NASA-TLX suggests that drivers perceived the 2-back condition as more demanding than the 0-back condition. The mental and temporal demands were higher for the 2-back condition. Moreover, participants felt more effort was required to complete the 2-back condition and were more frustrated by the 2-back condition compared to the 0-back condition. These results were anticipated. However, none of the previous studies employing the speed regulation n-back task [9, 16, 18] reported scores of NASA-TLX or other comparable measures.

Moreover, this study was the first attempt to use an objective measure of MW, namely the mean P3a amplitude evoked by novel environmental sounds, to assess the speed regulation n-back task. The P3a results suggest that the different levels of the n-back task had a significant effect on the amount of resources used to process the novel environmental sounds. Employing the multiple resources theory [21], we argue that the 2-back condition required more processing resources; therefore, fewer resources were left to process the novel environmental sound. This effect can be observed as a decreased mean P3a amplitude in the 2-back condition compared to the 0-back condition. This effect was anticipated based on the literature review.

The differences in subjective MW between the two conditions suggest a large effect ($d = 1.528$) for the overall score of NASA-TLX. Hence, we argue that there were substantial differences in demands between the two conditions and, therefore, the changes in the P3a amplitude can be explained in the light of the multiple resources theory [21] and resource competition [2]. However, we suggest that in the case of low-demanding tasks, researchers could potentially observe a floor effect. This could be due to the low demands of the task, which does not fully occupy the processing resources. As plenty of free resources are available to process the auditory cues, the P3a amplitude might not be in a direct reversed relationship with the demands of the primary task, and could therefore be non-sensitive to the demands of the primary task.

We observed an inversed relationship between the task demands and the P3a amplitude. This finding is in line with the majority of other studies (e.g., [5, 12, 14]). However, some authors argue that other factors might be modulating the P3a amplitude as well. The results of Cahn and Polich suggest that attentional withdrawal during meditation leads to a decreased P3a amplitude evoked by novel auditory stimuli. The results of Figalová et al. suggest a discrepancy between perceived MW and mean P3a amplitude evoked by novel auditory stimuli during automated driving in a realistic environment. These studies suggest that other factors, namely top-down attention control, novelty, environment, or arousal, might play an important role in modulating the P3a amplitude. These factors should be addressed methodologically in future studies.

The empirical data support both H1 and H2 and provide further empirical evidence for the reliability and validity of the speed regulation n-back task to manipulate MW in driving research. Therefore, the speed-regulation n-back task may be used to manipulate MW in a standardised way. Nevertheless, further methodological recommendations must be derived in order to ensure the results of

different studies conducted in different environments utilising different scenarios are directly comparable.

4.1 Limitations and Future Work

This study focused only on two levels of the n-back task (0-back and 2-back). For a fine-grained evaluation of the task, we recommend measuring more levels (e.g., 0, 1, 2, and 3 with each participant). Moreover, we recommend validating the paradigm on a general adult sample, as our sample consisted predominantly of young university students. Future studies should also address the effect of other variables, such as arousal or top-down attention control, on the auditory P3a amplitude.

4.2 Contribution and Novelty

This study evaluates the speed regulation n-back task, a novel approach to manipulating MW in driving research. We used methods of estimating MW induced by the speed regulation n-back task that have not been reported before in the context of the task (NASA-TLX and mean P3a amplitude). We combined self-reported and EEG-based estimates of mental workload that provide robust evidence of the validity and reliability of the task. Moreover, our empirical data help to comprehensively understand the cognitive mechanisms underlying the task.

CONCLUSION

The data suggest that drivers perceived the 2-back condition as more demanding than the 0-back condition. We found differences in the mental and temporal demand, effort, and frustration dimensions of NASA-TLX. Moreover, the P3a amplitude was significantly decreased for the 2-back condition compared to the 0-back condition. The speed regulation n-back task appears to be a reliable, valid, and reproducible method to manipulate MW in driving research.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement 860410. The project was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)-Project-ID 416228727-SFB 1410.

REFERENCES

- [1] B Rael Cahn and John Polich. 2009. Meditation (Vipassana) and the P3a event-related brain potential. *International journal of psychophysiology* 72, 1 (2009), 51–60.
- [2] Marvin M Chun, Julie D Golomb, and Nicholas B Turk-Browne. 2011. A taxonomy of external and internal attention. *Annual review of psychology* 62 (2011), 73–101.
- [3] Dick De Waard and KA Brookhuis. 1996. The measurement of drivers' mental workload. (1996).
- [4] David Embrey, Claire Blackett, Philip Marsden, and Jim Peachey. 2006. Development of a human cognitive workload assessment tool. *MCA Final Report, Lancashire* (2006), 1–253.
- [5] Carles Escera, Kimmo Alho, István Winkler, and Risto Näätänen. 1998. Neural mechanisms of involuntary attention to acoustic novelty and change. *Journal of cognitive neuroscience* 10, 5 (1998), 590–604.
- [6] Monica Fabiani, Victoria A Kazmerski, Yael M Cycowicz, and David Friedman. 1996. Naming norms for brief environmental sounds: Effects of age and dementia. *Psychophysiology* 33, 4 (1996), 462–475.
- [7] Nikol Figalová, Hans-Joachim Bieg, Julian Elias Reiser, Yuan-Cheng Liu, Martin Baumann, Lewis Chuang, and Olga Pollatos. 2023. From Driver to Supervisor: Comparing Cognitive Load and EEG-based Attention Allocation across Automation Levels. *arXiv preprint arXiv:2306.08477* (2023).
- [8] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
- [9] Moritz Held, Jochem W Rieger, and Jelmer P Borst. 2022. Multitasking while driving: central bottleneck or problem state interference? *Human factors* (2022), 00187208221143857.
- [10] Philipp Hock, Johannes Kraus, Franziska Babel, Marcel Walch, Enrico Rukzio, and Martin Baumann. 2018. How to design valid simulator studies for investigating user experience in automated driving: review and hands-on considerations. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 105–117.
- [11] Marius Klug, Sein Jeung, Anna Wunderlich, Lukas Gehrke, Janna Protzak, Zakaria Djebbara, Andreas Argubi-Wollesen, Bettina Wollesen, and Klaus Gramann. 2022. The BeMoBIL Pipeline for automated analyses of multimodal mobile brain and body imaging data. *bioRxiv* (2022).
- [12] Arthur F Kramer, Leonard J Trejo, and Darryl Humphrey. 1995. Assessment of mental workload with task-irrelevant auditory probes. *Biological psychology* 40, 1-2 (1995), 83–100.
- [13] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. 2019. ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage* 198 (2019), 181–197.
- [14] John Polich. 2007. Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology* 118, 10 (2007), 2128–2148.
- [15] Menja Scheer, Heinrich H Bülthoff, and Lewis L Chuang. 2016. Steering demands diminish the early-P3, late-P3 and RON components of the event-related potential of task-irrelevant environmental sounds. *Frontiers in human neuroscience* 10 (2016), 73.
- [16] Jakob Scheunemann, Anirudh Unni, Klas Ihme, Meike Jipp, and Jochem W Rieger. 2019. Demonstrating brain-level interactions between visuospatial attentional demands and working memory load while driving using functional near-infrared spectroscopy. *Frontiers in human neuroscience* (2019), 542.
- [17] Evona Teh, Samantha Jamson, Oliver Carsten, and Hamish Jamson. 2014. Temporal fluctuations in driving demand: The effect of traffic complexity on subjective measures of workload and driving performance. *Transportation research part F: traffic psychology and behaviour* 22 (2014), 207–217.
- [18] Anirudh Unni, Klas Ihme, Meike Jipp, and Jochem W Rieger. 2017. Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in human neuroscience* 11 (2017), 167.
- [19] Remo MA Van der Heiden, Christian P Janssen, Stella F Donker, Lotte ES Harde- man, Keri Mans, and J Leon Kenemans. 2018. Susceptibility to audio signals during autonomous driving. *PloS one* 13, 8 (2018), e0201963.
- [20] Anne E Wester, Koen BE Böcker, Edmund R Volkerts, Joris C Verster, and J Leon Kenemans. 2008. Event-related potentials and secondary task performance during simulated driving. *Accident Analysis & Prevention* 40, 1 (2008), 1–7.
- [21] Christopher D Wickens. 2008. Multiple resources and mental workload. *Human factors* 50, 3 (2008), 449–455.