Analysing Behavioural and Eye-Tracking Data to Investigate Working Memory Load and Visuospatial Demands During Driving

Bachelor's Project Thesis

Nilma P. Kelapanda, s3149323, n.p.kelapanda@student.rug.nl,
Supervisors: Dr J. P. Borst & M. Held, M.Sc

Abstract: The incorporation of systems assisting the driver of a car has gained interest in the past decade. Driving is a complex task, deeming research into adaptive automation, in which control of the car is split between the driver and a built-in system that takes over when necessary, a valuable study. For such a system, reliable measures of cognitive load, defined in this study as working memory load (WML) and visuospatial demands (VD), need to be determined in order to notify the system when the driver may need assistance. This study focused on analysing behavioural and eye-tracking data of participants in a simulated driving environment (highway) in order to assess cognitive load and the effects on working memory performance and driving performance. The four research questions explored were the following; *Does cognitive load have an influence on working memory performance?*; *Does cognitive load have an influence on driving performance?*; *Can pupillometry predict cognitive load while driving?*; *Does cognitive load have an influence on the frequency of speedometer checking?* WML was manipulated by a speed regulation version of the n-back task ($n = 0, 1, 2, 3, 4$). VD was manipulated through a construction site with narrower and fewer lanes. Results indicated a significant decrease in working memory performance and frequency of speedometer checking as WML increased. Alongside this, a significant decrease in driving performance and increase in driving difficulty as VD increased was observed. Finally, it was found that pupil size was a predictor for WML. Findings of this study can be applied to adaptive automation as measures of cognitive load and predictors of working memory performance have been significantly determined.

1 Introduction

The task of driving is not a simple one. Controlling a car in general traffic conditions is divided into several sub-tasks which require a combination of cognitive demands. Information from road signs, external traffic participants, and internal meters/displays must all be dynamically integrated and updated in order to drive safely. This integration and combination of demands puts a high load on the driver's working memory.

Working memory is an actively investigated topic in the field of psychology and refers to the items of information that are held in the memory as cognitive tasks are executed (Cowan, 2014). A driver's ability to successfully execute the act of driving is influenced by their working memory load. External tasks could result in a higher load, meaning less of the driver's resources are available to focus on driving (Nijboer, Borst, van Rijn, and Taatgen, 2016).

An accurate measure of working memory load could lead to the incorporation of systems within a vehicle aiming to assist the driver with the cognitive load that comes with operating the vehicle. Over the past decade, the interest in such adaptive driving, in which the control is dynamically divided between the driver and the vehicle, has increased. *Adaptive automation*, conceived by (Parasuraman, Cosenzo, and de Visser, 2009) is one such idea for the integration of a system that automates driving by adapting to the driver's current state of working memory load. For an effective implementation of this, working memory load needs to be reliably measured.

Unni et al. (2017), looked into corresponding brain areas during a driving task in order to predict variations in working memory levels. This was done through means of a modified n-back task involving speed signs, in a simulated virtual driving experiment. Results showed changes in driving behaviour, such as an increase in accelerator variability, as the working memory load increased due to the speed regulation task.

Scheunemann et al. (2019) looked into predicting visual attention, in terms of driving performance, independent of n-back level. For this, steering reversal rate, that is, how often a driver crosses the center position of the steering wheel, was analysed as an indicator of driving performance. Higher working memory loads had the effect of increased steering reversal rates. An interaction between visual attention and working memory level was also observed; the effect of a more difficult driving situation caused a much larger drop in working memory performance for higher working memory levels.

To investigate the demands on working memory load during driving, this project will look into replicating the behavioural driving results of Unni et al. under similar conditions. The effects of visuospatial conditions will also be investigated as a measure of driving difficulty. Participants will drive on a simulated highway while doing the n-back task in the form of memorising speed signs and driving according to the speed limit of n speed signs ago. The visuospatial attention will be manipulated by the implementation of a construction site on the highway with narrower lanes, giving us further insight into the effect on driving performance.

This study will focus on answering the following research questions, split into two main interests: behavioural and eye-tracking. Concerning the investigation into behavioural data (driving behaviour), the research questions are "*Does cognitive load have an influence on working memory performance?*", and "*Does cognitive load have an influence on driving performance?*". With these questions, the study aims to explore what effect cognitive load (defined as the working memory load and the visuospatial demands) has on both the working memory performance and the driving performance.

Secondly, concerning the eye-tracking data, the research questions are "*Can pupillometry predict cognitive load while driving?*" and "*Does cognitive load have an influence on the frequency of speedometer checking?*". The first question deals with using the size of pupil dilation to predict cognitive load. The second question will explore the effect of cognitive load on how often the speedometer is checked during driving. Frequency of speedometer checking could give us insight into the priorities of the driver in different cognitive load settings. The task is to regulate speed, meaning that the driver should be keeping an eye on the speed through the speedometer. The frequency of this may change as cognitive load varies.

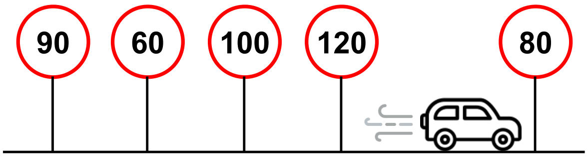
The hypothesis of this study in the behavioural component is that an increase in working memory load through means of an n-back task and a variation of visuospatial conditions (increased difficulty) will correlate with a decrease in working memory performance and driving performance, measured through an increase in speed error, steering reversal rate, lane deviation, and number of collisions with other traffic. This follows from the findings by Scheunemann et al. (2019) and Unni et al. (2017). Regarding eye-tracking, pupil size is predicted to increase as n-back increases, with the possibility of a drop in size if the participant gives up in the second half of the 4-back task. Finally, frequency of speedometer checking is expected to have a negative correlation with the n-back task and the increased visuospatial demands, as less control updates are expected to be performed with higher cognitive load (Salvucci and Taatgen, 2011).

2 Methods

Following Unni et al. (2017) an n-back task with five levels ($n = 0, 1, 2, 3, 4$) was used to manipulate working memory load. This task was integrated into the driving task by means of speed regulation. The designed environment also manipulated the visuospatial demands through a construction site with narrower lanes.

2.1 Participants

A total of 38 volunteers (23 male, 12 female, 3 other) aged 20-36 ($M = 23.1 \pm 3.0$), possessing a standard European driver's license, participated in this experiment. The participants, on average, obtained their driver's license 4.5 (± 3.1) years ago. All participants signed an informed consent form prior to the experiment and were compensated €12 for their participation.



	Current speed (km/h)	Sequence to be remembered & driven for subsequent speed signs
0-back	80	-
1-back	120	80
2-back	100	120, 80
3-back	60	100, 120, 80
4-back	90	60, 100, 120, 80

Figure 2.1: Example of the n-back experimental paradigm to manipulate cognitive workload. Consider a scenario where the participant is about to pass the 80km/h speed sign and the previous four speed signs were as shown in the schematic. For the corresponding n-back task, participants had to memorize the last n speed signs and drive at the n^{th} speed sign which occurred previously. For example, at 1-back, the participant's target speed is the previous sign (120 km/h) and has to keep the current speed sign in memory (80 km/h). Figure adapted from Unni et al. (2017), caption taken and slightly adapted from Scheunemann et al. (2019).

2.2 Experimental Set-up

The experiment took place on a simulated, straight, three-lane highway (see appendix appendix A.1). The features of the environment were minimal. Either side of the road was coloured green, representing grass. There were no median strips dividing the road from the rest of the environment. Apart from a single other car, represented by a blue rectangle and referred to as the *autocar*, there were no other objects/traffic on the highway. The autocar would stick to traffic rules such as overtaking from the left, staying on the right lane as much as possible, and following the current speed limit.

A dashboard was placed at the bottom, containing the speedometer of the car (as an integer). When the left or right indicators were pressed, they would appear on the respective sides of the dashboard as orange blinking arrows. The simulation had three rear-view mirrors: one on the top, one on the left, and one on the right. The autocar was visible in the corresponding mirror when it was behind the participant's car.

There were two experimental settings: normal and construction-site. The normal driving condition had lanes that were 3.5 meters wide, modeled after national German highway lane widths. The participant's car was 1.5 meters wide, meaning if centered, there was a meter of free space on the lane on each side of the car. In the construction condition (see appendix A.2), the width of the lanes were reduced by a meter to 2.5 meters. The lanes were separated by a full yellow line and the leftmost lane was closed off by a continuous row of pylons.

Speed signs that passed were identical to general speed signs in The Netherlands; black digits enclosed by a red circle (see appendices A.1 and A.2).

Within each trial the participants were presented with at least nine speed signs. The first speed sign appeared after 5 seconds, with each following speed sign appearing at intervals of 20 seconds. For n-back tasks with $n \geq 1$, there was a build-up phase of n speed signs preceding the nine speed signs where the participant would perform the task. For example, for $n = 4$, the build-up phase would be the first four speed signs. After the build-up phase, the task of speed regulation would start. Due to a difference in length of build-up phases per n-back trial, each trial differed in total number of speed signs, as well as time taken. An illustrative outline is presented in figure 2.1.

Participants interacted with the simulation using a steering wheel with blinkers, and a throttle and brake pedal (Driving Force GT by Logitech). The steering wheel was secured to the table in front of the screen and remained in the same location for all participants. The pedals were placed on the floor such that participants could move it closer or further depending on their level of comfort. An eye-tracking camera (EyeLink Portable Duo by SR Research), placed between the screen and the steering wheel, was used to continuously record the eye movements and pupil size of participants.

2.3 Experimental Procedure

The procedure of the current experiment closely follows that of Scheunemann et al. (2019). A trial consisted of 9 speed signs after a build-up phase of n speed signs and lasted around three minutes. The experiment consisted of 20 trials in total, divided by a short break into two blocks of 10 trials each. The entire experiment took about 60 minutes. Within a block, each n-back trial ($n = 0, 1, 2, 3, 4$) appeared twice: once with a construction site and once without. The order of the trials was determined pseudorandomly with a few constraints. Firstly, no n-back level could appear twice in a row. Secondly, the construction/non-construction conditions were alternated from trial to trial. These constraints on the randomization were incorporated with the aim of avoiding habituation effects for the memory task and the visuospatial demands. Finally, the order of the trials in the first block was reversed to form the order of trials in the second block.

Prior to performing the experiment the participant was given instructions about the driving and the memory task. They then performed a practice round (one 2-back trial on the normal highway and a total of 5 speed signs) to get accustomed to the simulation and the steering wheel. Next, the eye-tracker was calibrated. This involved the participant following a target around the computer screen with their eyes. This procedure was repeated twice: once to calibrate and once to validate whether that calibration was accurate. Calibration was performed again in case the validation was inaccurate.

After calibration, the experiment began. Each trial was preceded by a pop-up message appearing, telling the participant which n-back task they should perform in the coming trial (see appendix A.3). The percentage of total trials they had already completed was also shown in the message. The trial would only begin after the participant clicked the X button on the steering wheel indicating they were ready for the trail to start. After this the data would start recording. Furthermore, every trial (excluding the very first one) was preceded by an eye-tracking drift correction. This required the participant to look at a target at the center of the screen. If the measured eye position deviated too far from the position of the target, calibration was performed again. Otherwise the deviation was automatically taken into account with recording of the eye position.

2.4 Data collection

Behavioral data was recorded to track the participant's driving behavior and performance on the n-back task. The raw variables were recorded every 5 milliseconds. To assess the participant's driving behavior, the steering angle of the steering wheel was recorded. The position and orientation of the participant's car and the autocar was recorded as well. The car's position is used to determine driving performance in terms of lane centering. The speed of the car was also recorded and used to determine the error rate of the speed regulation task. The moment when a speed sign appears was also recorded, which is useful to determine when the participant is expected to change his/her speed. Lastly, a variable was recorded that tracked when the blinkers were used. This helps determine when a lane change was initiated and how long it took.

The eye-tracker records a number of raw variables at a rate of 500 Hz. Eye positions were measured in x and y coordinates relative to the PC monitor (1920×1080 px). Pupil dilation is measured in terms of diameter, in arbitrary units. As these units differ for all participants, raw pupil dilation cannot be compared across participants. Instead, it must be compared to some base-level dilation for each individual participant, explained in the following sub-section of this paper.

In addition to recording the raw data the eye-tracker software automatically sorts the data into fixations, saccades and blinks, removing a significant amount of noise.

2.5 Data Analysis

The overall analyses for both the driving behaviour and eye-tracking analysis was done in the programming language R (R Core Team, 2020). The build-up phase was excluded from all analyses.

For the working memory performance, error rates in the speed regulation task were calculated by looking at the percentage of speed error of the target speeds for each participant. This was done manually by looking each speed sign after the build up phase for each participant and checking if the target speed was reached and maintained for enough time. The following formula mathematically expresses how this was done:

$$\text{error rate} = \frac{\#\text{ incorrect target speeds in trial}}{\#\text{ speed signs in trial}(=9)}$$

For the driving performance, steering reversal rate, number of collisions, and lane deviation was calculated. Steering reversal rate was calculated by the following formula, where a steering reversal is defined as crossing the center of the steering wheel:

$$\text{steering reversal rate} = \frac{\text{total steering reversals}}{\text{seconds in a trial}(=165)}$$

Number of collisions was calculated by the following formula, where number of interactions refers to interactions between the participant's car and the other car in the simulation. Interactions are defined by the participant's car attempting to overtake the other car.

$$\text{proportion collisions} = \frac{\#\text{ collisions}}{\#\text{ interactions}}$$

Lane deviation was calculated by taking the absolute deviation from the center of the lane (per lane). This was done by taking the participant's car position (x-axis value) and first determining which of the three lanes it was in. Then, the lane deviation was recorded as the absolute distance from the center of whichever lane the participant was in. Lane-changing manoeuvres were excluded from this calculation, by removing the last 1.5 and the next 1.5 data points from the analysis once the center of the participant's car crossed over into another lane. This 3-second window was chosen as the estimated time taken for a lane-change from start to finish. The mean lane deviation per lane was averaged for all separate lanes per trial, per visuospatial condition.

For pupil size measurements, a subtractive baseline correction was applied, resolving any fluctuations between trials (Mathôt et al., 2018). For this correction, a baseline must be chosen. the first five seconds of the experiment, from the start till the first speed sign appeared, was selected for the baseline. A mathematical formula for this method is as follows, in arbitrary units:

$$\begin{aligned}&\text{corrected pupil size}\\ &\qquad = \text{pupil size} - \text{baseline pupil size}\end{aligned}$$

For the calculating of frequency of speedometer checking, the area of the speedometer fixation differed so we first isolated an area that was all-encompassing for the fixations of all participants. This was selected manually. The formula used for this calculation is the following:

$$\begin{aligned}&\text{frequency of speedometer checking}\\ &\qquad = \frac{\#\text{ fixations on speedometer in a trial}}{\text{total fixations in a trial}}\end{aligned}$$

Before starting the analysis, the participants' data was evaluated to see if any data should be excluded. Firstly, 11 participants were excluded due to an incorrect trial length, deeming the data inaccurate for analysis. Five more participants were excluded because they performed poorly in the 0-back and 1-back, suggesting either that they were not taking the experiment seriously, or did not understand the n-back task. This means that there was a total of 16 participants that were excluded from all analyses. For the analysis of error rate in the speed regulation task and the lane deviation, no additional participants were excluded, resulting in a sample group of $n_p = 22$. For the analysis into the number of collisions, three more participants were excluded, reducing the sample group to $n_p = 19$. Due to a lack of steering data, an additional 16 participants were excluded from the steering reversal analysis, resulting in a sample group of $n_p = 7$. Finally, for the eye-tracking analysis, six participants alongside the original exclusion of 16 participants, had unusable data, putting the sample group to $n_p = 16$.

3 Results

This section is split into two parts: *Driving Behaviour*, and *Eye-Tracking*. This was done for ease of understanding which of the two sets of research questions we aim to answer with the results.

3.1 Driving Behaviour

3.1.1 Error Rate in Speed Regulation Task

In figure 3.1 the error rates in the n-back speed regulation are presented. In the 0-back level, we see that there is little error in the task for both visuospatial conditions. The 1-back level has low error

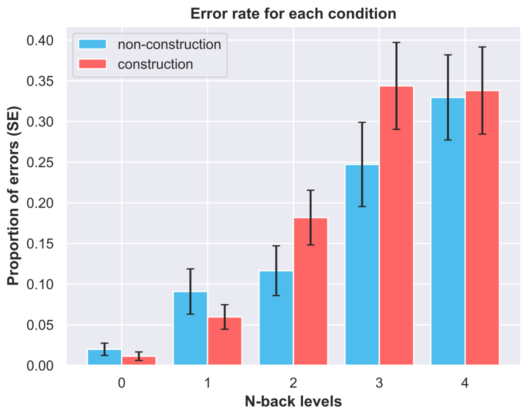


Figure 3.1: Error rate in the speed regulation task, per n-back level, per visuospatial condition, including bars representing the standard error from the mean ($n_p = 22$). Mooij (2021)

n-back level	p-value	Bonferroni
0-back – 1-back	*$3.80e^{-4}$	*$4.00e^{-3}$
0-back – 2-back	*$2.52e^{-7}$	*$2.52e^{-6}$
0-back – 3-back	*$3.04e^{-11}$	*$3.04e^{-10}$
0-back – 4-back	*$3.04e^{-11}$	*$3.04e^{-10}$
1-back – 2-back	*0.006	0.062
1-back – 3-back	*$1.06e^{-6}$	*$1.06e^{-5}$
1-back – 4-back	*$1.01e^{-8}$	*$1.01e^{-7}$
2-back – 3-back	*$1.35e^{-5}$	*$1.35e^{-4}$
2-back – 4-back	*$4.27e^{-7}$	*4.27^{-6}
3-back – 4-back	0.311	1.00

Table 3.1: Results of the paired sample t-test. Significant results are marked with an asterisk

rates for both visuospatial conditions. For the 2-back level, the non-construction error rate is very similar to that of the 1-back level, whereas the construction condition has a much higher error rate. The error rates for the 3-back level are higher than the 2-back level, with the construction condition being higher than the non-construction condition. Finally, the 4-back level error rates are very similar between visuospatial conditions, however remain lower than the construction condition of the 3-back level.

The significance of the results in figure 3.1 were analysed with a two-way repeated measures ANOVA. The factors used for this test were the n-back level and the visuospatial condition. It was found that n-back level had a significant effect $[F(4, 84) = 26.46, p < 0.001]$ on the error rate, whereas visuospatial condition had no significant effect. The test also indicated a marginal interaction effect between n-back task and visuospatial condition $[F(4, 84) = 2.51, p = 0.048]$.

As there were many n-back levels, we further investigate the effect on error rate through use of a paired sample t-test. Results of this test are presented in table 3.1, where each pair of n-back level is listed with the corresponding p-value and *Bonferroni* correction to the p-value. The table shows that almost all pairs of n-back levels have a significant difference in mean error rate. The exceptions are the 1-back–2-back pair, and the 3-back–4-back pair. This implies that there exists a grouping effect of the n-back levels into three groups: the lower n-back level consisting of the 0-back task, the medium n-back level consisting of the 1-back and the 2-back tasks, and the higher n-back level consisting of the 3-back and 4-back tasks.

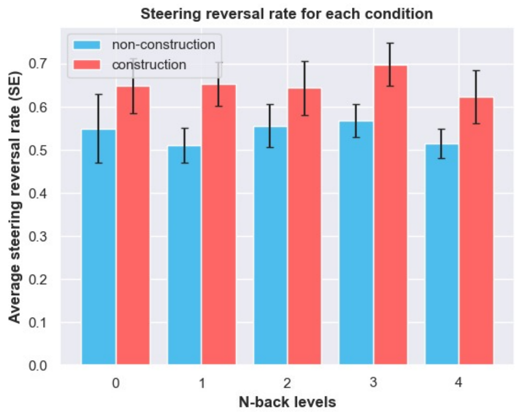


Figure 3.2: Steering reversal rate per n-back level per visuospatial condition, including bars representing the standard error from the mean ($n_p = 7$).

3.1.2 Steering Reversal Rate

The steering reversal rate is presented in figure 3.2. Across all n-back conditions in figure 3.2, the non-construction condition appears to have no significant difference, whereas the construction condition has higher rates that are also similar to one another.

After running a two-way repeated measures

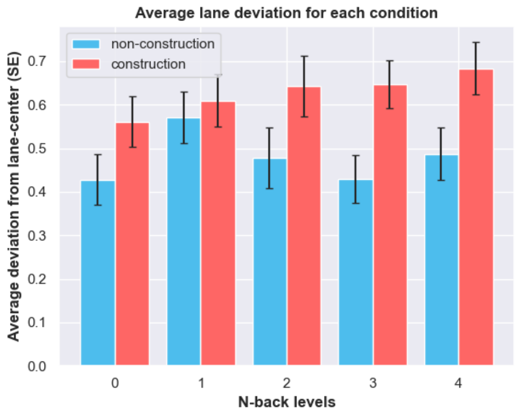


Figure 3.3: Lane deviation (in distance units used in the simulation, referring to meters) from the center of the lane per n-back level per visuospatial condition, including bars representing the standard error from the mean ($n_p = 22$).

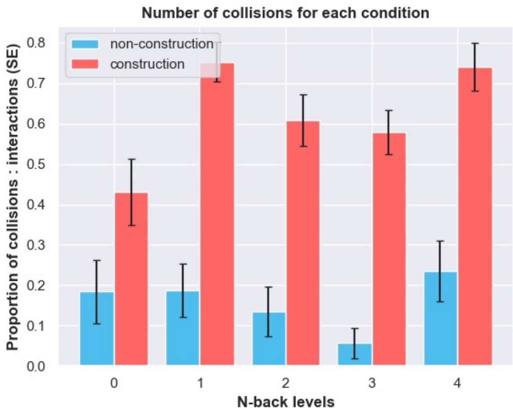


Figure 3.4: Average number of collisions as a ratio to the number of interactions per n-back level, per visuospatial condition, including bars representing the standard error from the mean ($n_p = 19$). Mooij (2021)

ANOVA on the factors n-back level and the visuospatial conditions, it was found that the n-back level had no significant effect on steering reversal rate. There was also no significant interaction between n-back level and visuospatial condition. On the other hand, there was indeed a significant effect of visuospatial condition on steering reversal rate $[F(1, 6) = 20.1, p < 0.001]$, which implies that the driving difficulty was increased when participants drove on the construction site as opposed to the non-construction site.

3.1.3 Lane Deviation

The average lane deviation is presented in figure 3.3. The figure indicates a steady increase in lane deviation as n-back level increases in the construction condition. For the non-construction condition, however, there is no clear effect of n-back level increase on lane deviation. The lane deviation is higher for each construction condition than the non-construction condition per n-back level. It should also be noted that the lane deviation in the 1-back non-construction condition has a value similar to that of the 0-back construction conditions.

After running a two-way repeated measures ANOVA with n-back level and visuospatial condition as factors, it was found that neither n-back level and interaction between n-back level and visuospatial condition had a significant effect on lane deviation. The visuospatial conditions did, however, have a significant effect on lane deviation $[F(1, 21) = 14.52, p < 0.005]$. As lane deviation is an indicator of driving performance, these results show that driving performance decreases when driving through the construction site in comparison to the non-construction site.

3.1.4 Collisions

The number of collisions is presented in figure 3.4. There is a noticeable difference in number of collisions between visuospatial conditions in the figure. There appears to be a possible decrease in number of collisions in the non-construction condition from the 0-back/1-back to the 3-back level, whereas the 4-back level has a higher number of collisions than the other n-back levels.

A two-way repeated measures ANOVA test was used to test the significance, with n-back level and visuospatial condition as factors. Results of this test showed no significant effect of n-back level or the interaction between n-back level and visuospatial condition, however did show a significant effect of visuospatial condition on the number of collisions $[F(1, 18) = 58.25, p < 0.001]$. This indicates that the driving difficulty, as measures by the number of collision in relation to the number of interac-

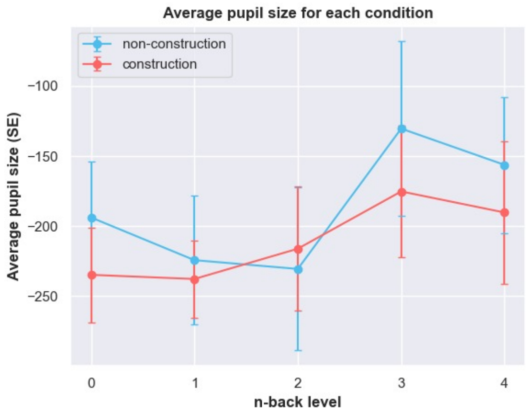


Figure 3.5: Average pupil size per n-back level (shown in arbitrary units), per visuospatial condition, including bars representing the standard error from the mean ($n_p = 16$). Lijnzaad (2021)

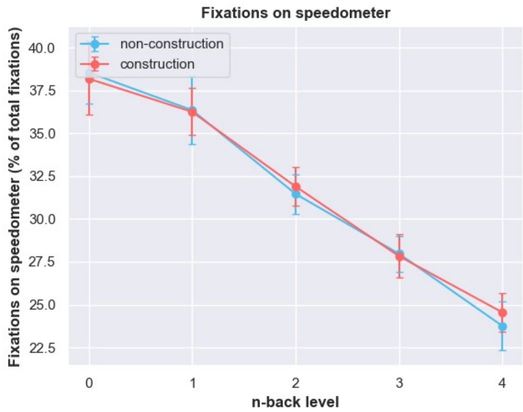


Figure 3.6: Average number of fixations on the speedometer as a percentage of total fixations per n-back level, per visuospatial condition, including bars representing the standard error from the mean ($n_p = 16$). Lijnzaad (2021)

tion between cars, increases when the participant drives in the construction site as opposed to the non-construction site.

3.2 Eye-Tracking

3.2.1 Pupil Size

The average pupil size is presented in figure 3.5. Initially, for the non-construction condition, the figure suggests that the average pupil size decreases linearly for the first three n-back levels, but then increases in the 3-back and 4-back levels. The pupil size during the construction condition appears to increase more steadily as n-back level increases. From the figure it can also be seen that there the pupil size is generally lower for the construction condition than for the non-construction condition, where the 2-back level is an outlier.

The results of a two-way repeated measures ANOVA test on the factors n-back level and visuospatial condition showed that there was no significant effect of visuospatial condition or interaction between n-back level and visuospatial condition on pupil size. However, the effect of n-back level on pupil size was significant $[F(4, 60) = 2.97, p < 0.05]$.

3.2.2 Fixations on the Speedometer

The average number of eye fixations on the speedometer is presented in figure 3.6. In this figure we can see almost no difference in number of fixations between visuospatial conditions. Alongside this, there appears to be a clear negative linear relationship between n-back level and the number of fixations on the speedometer.

After running a two-way repeated measures ANOVA test on these results with n-back level and visuospatial conditions as factors, it was found that there was no significant effect of visuospatial condition or interaction between n-back level and visuospatial condition on fixations to the speedometer. There was, however, an unsurprising significant effect of n-back task on number of fixations to the speedometer $[F(4, 60) = 47.86, p < 0.001]$.

4 Discussion

This research project was focused on answering multiple questions regarding both driving behaviour and eye-tracking data, with the aim of analysing working memory load (WML) and visuospatial demands (VD). After collecting and testing the results of the experiment, we move onto discussing the relevance of the findings and the implications that follow. This will be done by splitting the discussion into the respective analyses (driving behaviour and eye-tracking) with conclusions, followed by an exploration of the implications of the findings.

4.1 Driving Behaviour

The first research question that shaped the driving behavioural portion of this study was "*Does cognitive load have an influence on working memory performance?*". This was investigated by analysing the error rate in the speed regulation task. A significant increase in errors were observed as n-back level increased. In terms of cognitive load, these results indicate a decrease in working memory performance when WML was increased, but not when VD were increased. The hypothesis that both WML and VD would influence working memory performance is therefore deemed partially false. VD having no significant effect on working memory performance was counter-intuitive considering the study conducted by Scheunemann et al. (2019), which asserted that an increase in VD could result in a decrease in working memory performance.

The next research question regarding driving behavioural analysis was "*Does cognitive load have an influence on driving performance?*". This was investigated by analysing the following frequency measures, as shown by McLean and Hoffmann (1975) to be useful indicators of driving performance: steering reversal rate per second and number of collisions with respect to number of interactions. A significant increase in steering reversal rate, lane deviation, and number of collisions was observed as VD increased, that is, participants traversed the construction site. This indicates that drivers in more visuospatially demanding settings experience an increase in driving difficulty, reducing their driving performance. Taking these findings into consideration, the hypothesis that both WML and VD would influence working memory performance is therefore deemed partially false. WML having no significant effect on working memory performance was counter-intuitive considering the study conducted by Scheunemann et al. (2019), which asserted that an increase in WML could result in a decrease in driving performance.

4.2 Eye-Tracking

The eye-tracking portion of this study aimed to answer another two research questions, the first one being "*Can pupillometry predict cognitive load while driving?*". This was investigated by measuring pupil size changes during a trial through means of an eye-tracker. A significant increase in pupil size was observed as n-back level increased. In terms of cognitive load, these results imply that an increase in pupil size can be seen as a predictor for the WML component of cognitive load while driving. This is in line with the hypothesis corresponding to this research question.

The final question this study aimed to explore was "*Does cognitive load have an influence on frequency of speedometer checking?*". This was investigated through use of an eye-tracker that recorded the saccades to the speedometer. A significant decrease in frequency of speedometer checking was observed as n-back level increased, that is, as the WML component of cognitive load increased. This is in line with the hypothesis corresponding to this research question.

4.3 Implications & Limitations

Overall, this study can conclude that working memory performance is influenced by WML, driving performance is influenced by VD, pupillometry is a predictor for WML, and frequency of speedometer checking is influenced by WML. Although some of these conclusions were expected, others clash with previous findings.

As mentioned earlier, Scheunemann et al. (2019) found WML to have an effect on driving performance and VD to have an effect on working memory performance. An interaction effect between WML and VD was also observed in their findings. Namely, an increase in cognitive demands in terms of WML resulted in a decrease in driving performance in terms of VD (Scheunemann et al., 2019). There was a change in activation patterns in the working memory level, caused by changes in driving difficulty (VD).

Such interaction observations could not be extracted from the results of the current study. None of the analyses found a significant interaction effect between WML and VD. One reason this may differ in the findings of this study could be attributed the difference in experimental design. Scheunemann et al. conducted the experiment with a total of 9 speed signs for each level of n, excluding a build-up phase. The experiment used for this study, on the other hand, incorporated a build-up phase. This resulted in a total of 9 speed signs for the analysis of each n-back task. It would be intuitive to expect the interaction effect to therefore be emphasised in this study, following from Scheunemann et al.. Nonetheless, the contrast in findings suggests some other distinction between experimental set-ups or some limitation of the current study to have resulted in no such interaction effect being observed.

The key limitations that presented itself in this study will now be taken into consideration. Firstly, the consistency of number of eligible participant data (n_p) for each analysis was lacking. The sample groups n_p differed greatly per analysis, with the lowest value being $n_p = 7$. With regard to the findings of Scheunemann et al., some sample groups may have been too low to observe significant (interaction) effects. Nevertheless, it may be the case that these conclusions were on the right track, deeming a replication of this study with more data to be an important next step in the investigation into adaptive automation.

Another substantial limitation of this study is the simulated driving environment used in the experiment. The environment was extremely minimal, with very little traffic or external/internal components implemented into it. This means that it could not properly imitate the cognitive task of driving in real life traffic conditions. For this reason, it may be difficult to translate the findings and conclusion into real-life applications.

In light of the initial problem of finding measures of cognitive load that can be absorbed into the application of adaptive automation (AA) in driving, we have found some valuable results. One such result was that driving performance is decreased as VD increase. This could be taken into account when designing such an AA. The AA could calculate the levels of VD by checking steering corrections or deviation from the center of the lane, and in turn, take over (partial) control of the car when it appears to take too much of a toll on the driver. Number of collisions is not measurable as a variable that the AA could take into account, however.

Furthermore, such an AA could be implemented alongside an eye-tracker. As results indicate, pupil size and frequency of speedometer checking are predictors of an increased WML. Pupil size could be tracked and corrected for with a similar subtractive baseline as performed in this study. Once changes in pupil size are observed, the AA would be able to predict an increase in WML and hence would be prepared to step in if driving performance begins to decrease as well. Or in the least, would be able to assist with some cognitive demands of the situation so as to take some load off of the driver. Similarly, a measured decrease in speedometer checking means the AA could display a reminder to the driver to divert attention back to the task of driving, as it indicates focus is being allocated to external tasks. These were just some possible implementations through which the significant results obtained in this study may be applied to the conception of semi-'self-driving' cars of the future, in the form of adaptive automation.

References

Nelson Cowan. Working memory underpins cognitive development, learning, and education. *Educational Psychology Review*, 26(2):197–223, 2014. ISSN 1040726X, 1573336X. URL `http://www.jstor.org/stable/43549792`.

G. Lijnzaad. *Predicting Working Memory Load and Visuospatial Demands During Driving Using Eye-Tracking.* unpublished bachelor thesis, University of Groningen, The Netherlands, 2021.

Sebastiaan Mathôt et al. Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods*, 50(1):94–106, February 2018. ISSN 1554-351X. doi: 10.3758/s13428-017-1007-2.

John R. McLean and Errol R. Hoffmann. Steering reversals as a measure of driver performance and steering task difficulty. *Human Factors*, 17(3):248–256, 1975. doi: 10.1177/001872087501700304. URL `https://doi.org/10.1177/001872087501700304`.

M. De Mooij. *Working Memory Load and Visuospatial Demands During Driving: A behavioral and eye-tracking analysis.* unpublished bachelor thesis, University of Groningen, The Netherlands, 2021.

Menno Nijboer, Jelmer P. Borst, Hedderik van Rijn, and Niels A. Taatgen. Driving and multitasking: The good, the bad, and the dangerous. *Frontiers in Psychology*, 7:1718, 2016. ISSN 1664-1078. doi: 10.3389/fpsyg.2016.01718. URL `https://www.frontiersin.org/article/10.3389/fpsyg.2016.01718`.

Raja Parasuraman, Keryl Cosenzo, and Ewart de Visser. Adaptive automation for human supervision of multiple uninhabited vehicles: Effects on change detection, situation awareness, and mental workload. *Military Psychology - MIL PSYCHOL*, 21:270–297, 04 2009. doi: 10.1080/08995600902768800.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

Dario D. Salvucci and Niels A. Taatgen. *The multitasking mind*. Oxford University Press, 2011.

Jakob Scheunemann, Anirudh Unni, Klas Ihme, Meike Jipp, and Jochem W. Rieger. Demonstrating brain-level interactions between visuospatial attentional demands and working memory load while driving using functional near-infrared spectroscopy. *Frontiers in Human Neuroscience*, 12, 2019. doi: 10.3389/fnhum.2018.00542.

Anirudh Unni, Klas Ihme, Meike Jipp, and Jochem W. Rieger. 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. doi: 10.3389/fnhum.2017.00167.

A Appendix

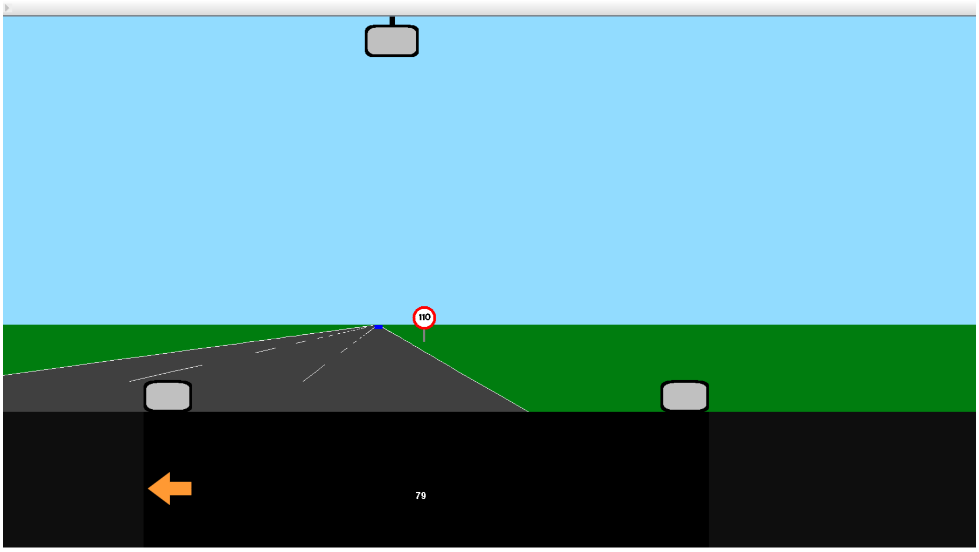


Figure A.1: Screenshot of the driving simulation in the non-construction condition. In the center of the black dashboard, the speedometer is printed as an integer in km/h. To the left of it is an example of the left indicator mid-blink. The rear-view mirrors (left, right, and top-middle) are present in the simulation, showing the autocar when it is behind the participant's car. The speed sign in this screenshot shows 110km/h.

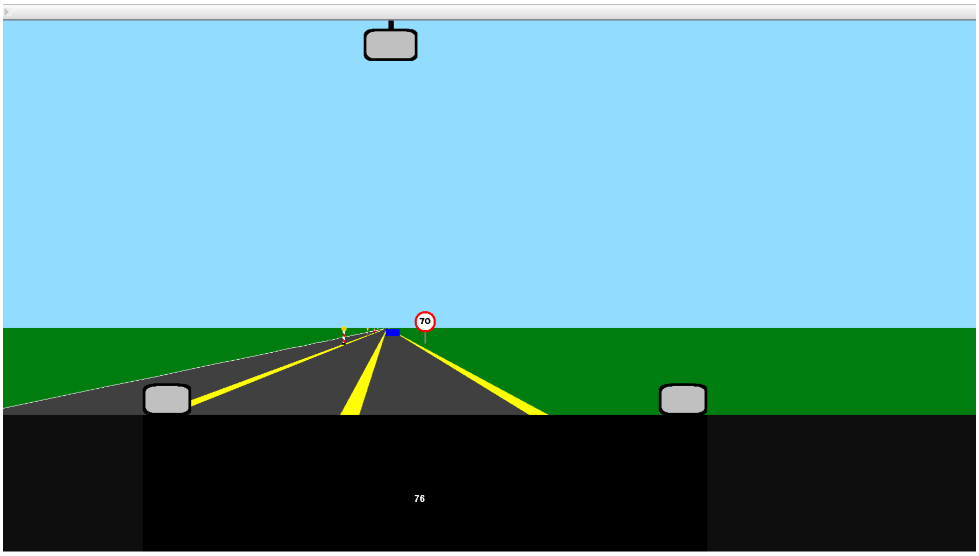


Figure A.2: Screenshot of the driving simulation in the construction condition. Identical to the non-construction condition except for the width of the lanes and the lines separating the lanes. In the same lane as the participant's car, the autocar (the blue rectangle) can be see in the distance. Pylons closing of the left-most lane are also visible. The speed sign in this screenshot shows 70km/h.

Figure A.3: Screenshot of the end non-construction condition, where the pop-up message has appeared before the next trial. The trial number, progress of the experiment (where a break is given at the 50% mark), and the n-back task for the next trial are all visible.