The Effect of the Number of Camera Feeds on Mental Workload and Monitoring Performance in Tunnel Traffic Control Room Operators

Léon D. Schlimme

Section Cognition, Data and Education, Department of Learning, Data-Analytics and Technology, University of Twente

201400213: Master thesis Human Factors & Engineering Psychology

Prof. Dr. Ing. W.B. Verwey

Dr. R.H.J. van der Lubbe

M.Sc. Nienke Bierhuizen

22.06.2025

Abstract

Previous research suggests that the number of closed-circuit television (CCTV) feeds monitored in tunnel traffic control rooms affects operator mental workload and performance. Research suggests a maximum of 12-16 feeds per operator, but operators often monitor more. This study investigated how feed count affects subjective and objective mental workload and performance. In a within-subjects design, 16 operators monitored prerecorded traffic across 8, 16, or 24 feeds, responding to incidents via keypress, signaling incident detection. Subjective mental workload was assessed via NASA Task Load Index (NASA-TLX), objective mental workload was measured via blink rate and pupil diameter, and performance was measured using response count, total response time and post-fixation response time. Data were analyzed using mixed effects models, with Karolinska Sleepiness Scale scores as a covariate. NASA-TLX scores, blink rates, response counts and response times did not differ with the number of camera feeds. Feed count influenced pupil diameter, with post-hoc tests unexpectedly showing larger pupil diameters at 8 feeds than 24, potentially due to screen brightness differences. Overall, results suggest that tunnel traffic control room operators may be able to monitor up to 24 feeds while maintaining good performance and experiencing no substantial changes in mental workload when only two distinct scenes are displayed. This highlights the importance of the complexity of visual information alongside feed count alone.

Introduction

Imagine sitting in front of a large video wall displaying many camera feeds from multiple tunnels. You are expected to monitor this large number of images and cannot miss anything of interest, while the number of feeds makes it difficult to monitor them all. Most of the time, nothing noteworthy happens, but any missed incident could result in preventable injury. This is a daily reality for operators monitoring road traffic in tunnels. While modern systems can help operators in automatically detecting some incidents, they are not able to detect all relevant incidents yet. The operator remains as a last line of defense and is expected to monitor relevant camera feeds and automated systems.

Sustaining attention across many feeds over extended periods of time is mentally demanding and exhausting. As the number of monitored feeds increases, mental workload increases as well. This study investigates how the number of monitored closed-circuit television (CCTV) feeds influences mental workload and operator performance in a monitoring task, a task frequently studied in vigilance research.

Vigilance

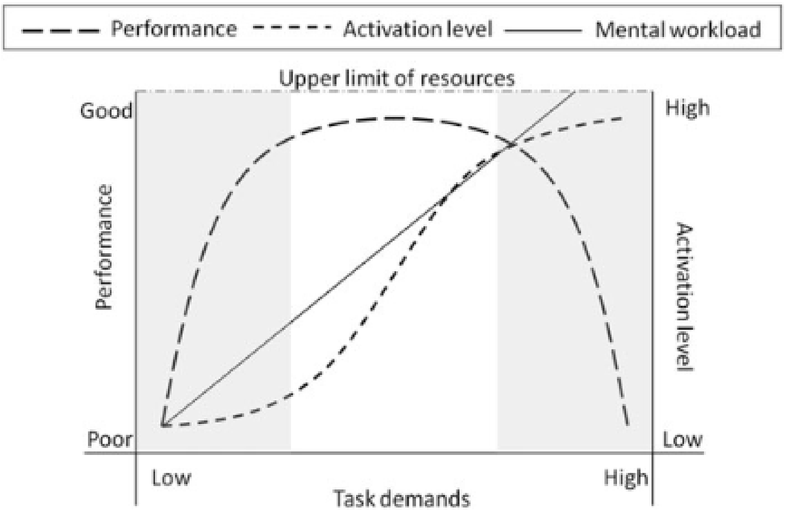
According to Davies and Parasuraman (1982), *vigilance* can be understood as the ability to sustain attention and to remain alert to stimuli over extended periods of time. This definition has been widely cited in vigilance literature (e.g. Warm et al., 2008) and fits the control room environment, which is why it was used in this thesis. Mackworth (1948) was one of the first to study the *vigilance decrement*, the finding that operator detection accuracy decreases over time, which sparked more research on the topic as a result. No explanation of the vigilance decrement is universally agreed upon, but an integrated perspective on explaining the phenomenon could offer the best explanatory power (Esterman & Rothlein, 2019).

Mental Workload

The present thesis used a definition of mental workload that is based on *attentional resource theory*. This theory posits that humans have limited attentional resources which can be allocated to multiple tasks (Kahneman, 1973). Mental workload was defined as follows: “The mental workload of a task represents the level of attentional resources required to meet both objective and subjective performance criteria, which may be mediated by task demands, external support, and past experience” (Young & Stanton, 2006, p. 818). Performance criteria can be imposed by external parties or represent one’s internal goals. Support can be provided by peers or technology. Past experience influences skills or knowledge, which in turn impact mental workload (Young & Stanton, 2004). Task demands are comprised of multiple task-related factors like time pressure and task complexity (Young et al., 2015). When task demands are increased beyond the available finite attentional resources, overload occurs, leading to performance decrements (Young & Stanton, 2004). When task demands are too low and cause insufficient stimulation, underload occurs. Available resources either shrink or are redirected for different use, resulting in performance decrements (Young et al., 2015). Consequently, building on the Yerkes-Dodson law (Yerkes & Dodson, 1908), attentional resource theory predicts maximal performance when task demands and mental workload are optimal (Hancock, 1989) (Figure 1).

Figure 1

Performance, Activation Level and Mental Workload as a Function of Task Demands



Note. From "State of science: mental workload in ergonomics" by M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, 2015, *Ergonomics*, *58*(1). Copyright 2015 by Taylor & Francis. https://doi.org/10.1080/00140139.2014.956151

Measuring Mental Workload

The most common measurements of mental workload include subjective ratings, physiological measures and performance measures (Young et al., 2015). For subjective scales, which assume that higher mental workload and a higher use of attentional resources lead to higher perceived effort, one popular option is the NASA-TLX (Hart & Staveland, 1988). Supporting the use of subjective scales, higher task demands have been linked to higher subjective mental workload scores (Warm et al., 1996; Warm et al., 2008). Physiological measures reflect arousal and are assumed to indicate objective mental workload. Popular examples include blink rate or pupil diameter (For a review, see Charles & Nixon, 2019). Supporting the use of these measures, higher task demands and effort have been associated with lower blink rates (Brookings, 1996; Stern et al., 1994; Veltman &

Gaillard, 1996) and larger pupil diameters (Biondi et al., 2023; Halverson et al., 2012; van der Wel & van Steenbergen, 2018). Performance measures assume that higher mental workload results in a decrease in performance if task demands exceed operator capacity (Young & Stanton, 2004). Common performance measures include response times and response counts or rates. The use of these measures is supported by findings showing that increasing task demands are associated with longer response times and lower detection rates (Helton & Russell, 2013; Helton & Warm, 2008).

Mental Workload and Human Performance

Mental workload has been shown to have an impact on several operator related factors besides performance. Excessive mental workload has been linked to increases in health risks (Klonowicz, 1995), burnout (Akca & Küçükoğlu, 2020; Zanabazar & Jigjiddorj, 2022), absenteeism (Jalali et al., 2023), fatigue, and decreasing job satisfaction (Rostami et al., 2021). To mitigate these negative consequences, it is important to understand and manage factors influencing mental workload.

One of these factors is task demands, which consist of task complexity and time pressure (Young et al., 2015). As mentioned above, higher task demands have been associated with higher objective and subjective mental workload. More specifically, in vigilance tasks, Warm et al. (1996) found subjective mental workload to increase with (1) a decrease in salience of critical signals, (2) an increase in spatial uncertainty of signal location and (3) an increase in event rate. These properties are especially relevant in CCTV-monitoring tasks. For instance, increasing the number of monitored CCTV feeds may may result in each feed occupying a smaller portion of the display, which could reduce the salience of critical signals as visual details in feeds become smaller. Additionally, monitoring a larger number of feeds may increase spatial uncertainty, as critical incidents can appear in a higher number of distinct feed locations. This also requires operators to divide their attention across more feed locations, increasing mental workload through higher attentional demands (Kahneman, 1973).

In addition to task demands, sleepiness is another relevant factor influencing operator mental workload, as their schedule requires them to work at different times of day, even night shifts. Fighting sleepiness as well as a lack of sleep have been linked to higher subjective mental workload (Verwey & Zaidel, 2000; von Gall et al., 2023). Additionally, higher sleepiness has been associated with changes in pupil diameter (Morad et al., 2009), higher spontaneous blink rate (Barbato et al., 2007) and other blink parameters such as blink duration or reopening time (Caffier et al., 2003).

Control Room Design and Mental Workload

While sleepiness may not be a factor easily addressed by changes in control room design, task demands are. As task demands are influenced by interface design, improved designs have been linked to reduced operator mental workload (MacDonald, 1999; Yan et al., 2017). For CCTV monitoring, higher image complexity and a higher number of monitored feeds have been associated with higher mental workload and lower detection rates (Gill et al., 2005; Neil et al., 2007; Pikaar, 2015; Warm et al., 1996). Past research has shown the maximum number of feeds that operators can monitor effectively while maintaining high detection rates to be 12-16 feeds per operator (Velastin, 2003, as cited in Schreibers et al., 2012; Wallace et al., 1997, as cited in Schreibers et al., 2012; Wood, 2007). Tickner and Poulton (1973), for instance, found detection rates of 83%, 84%, and only 64% when monitoring 4, 9, and 16 feeds in a security monitoring task, pointing toward serious decrements in detection performance for 16 feeds. These observations may be partly explained by the idea that scanning a higher number of feeds introduces a larger set size. Set size describes the number of items among which the target could appear (Lavie, 1995). According to perceptual load theory, perceptual load increases with set size, which leads to a higher use of attentional resources, which increases mental workload (Lavie, 1995). In addition to that, operators need more time to scan each tunnel when more feeds are presented per tunnel. This may increase time pressure, which has been linked to higher mental workload (Young et al., 2015). Importantly, introducing a higher number of feeds increases the number of information sources between which attention is divided. As the demands of each task between which attention is divided (in this case, the monitoring of each feed) drive performance (Kahneman, 1973) and mental workload, an increase in feeds increases mental workload, even if the demands for each feed are low.

In practice, control room operators frequently monitor a high number of feeds. In one of the Dutch tunnel traffic control rooms cooperating with vhp human performance, the company where this research was conducted, operators monitored up to six tunnel tubes simultaneously, with up to 16 feeds per tunnel tube. In a different study, Schreibers and Bouchier (2014) reported that even when monitoring more than 16 feeds, operators did not report overload. Instead, they perceived a large number of feeds as helpful for monitoring traffic. This disconnect between recommendations based on past research on one hand, and practical observations on the other hand may be explained by differences between control rooms in the way CCTV feeds are grouped into *scenes*. Scenes are defined as "… a logical and meaningful set of related and coherent images and other visual information, to be monitored with a specific aim" (Pikaar, 2015, p. 18). Because the monitored images in one scene usually share one subtask or geographical area, mental workload may be reduced. Each scene requires operators to maintain a separate mental model, which helps with the interpretation of information, and allows for a more efficient monitoring (Moray, 1999). Maintaining models requires working memory resources (Moray, 1999). As a result, mental workload increases with an increase in the number of monitored scenes, and with the number of feeds per scene (Pikaar, 2015). However, adding more feeds to one scene may also enable the perception of more elements of the situation. This may lead to improved situation awareness (Endsley & Garland, 2000) and the construction of a more complete mental model. This is especially relevant when incidents happen, which require adding additional feeds to the scene showing the incident in order to monitor more feeds of the incident area (Pikaar, 2015). If too few feeds are monitored, operators risk missing critical elements of a scene, which may negatively affect situation awareness and result in detection performance decrements. Mental workload may also increase due to the higher effort needed to integrate information from fewer feeds, especially when the geographical distance between feed locations is larger.

Prior research on CCTV monitoring mostly focused on the number of feeds, without considering the number of scenes shown by the feeds. However, as mentioned above, this distinction is important as monitoring a number of feeds belonging to the same scene may lead to lower mental workload than monitoring the same number of feeds across a higher number of scenes, as increasing the number of scenes increase task complexity and cognitive demands (Pikaar, 2015). In security control rooms, operators monitor a high number of separate scenes, each requiring a separate mental model (Wallace et al., 1997, as cited in Diffley & Wallace, 1998; Pikaar, 2015; Tickner & Poulton, 1973; Wood, 2007). In tunnel traffic control rooms, feeds are grouped into fewer scenes, with each tunnel tube representing one distinct scene (Pikaar, 2015). This may explain why traffic control room operators reported not feeling overload when monitoring higher numbers of feeds than recommended in past research, which was mostly conducted for security control rooms. As a result, past findings and recommendations on the ideal number of monitored CCTV feeds may not be applicable to tunnel traffic control rooms and should be reevaluated in this context.

Objective and Research Question

The present study aims to provide practical recommendations for making operator-centered design choices in tunnel traffic control rooms regarding the number of monitored feeds, specifically in contexts where two tubes are monitored, to prevent operator overload during vigilance monitoring tasks. To achieve this, it investigates how the number of camera feeds influences subjective and objective mental workload, as well as operator performance, while monitoring traffic.

Experiment and Hypotheses

The number of CCTV feeds (8, 16, or 24) was varied within participants as the independent variable in a simulated vigilance monitoring task. The feeds were evenly distributed between two tunnel tubes, requiring operators to maintain two mental models during monitoring for all experimental conditions. The number of feeds per condition was determined based on previous research (Wallace et al., 1997, as cited in Diffley & Wallace, 1998; Velastin, 2003, as cited in Schreibers et al., 2012; Wood, 2007), which indicated that the upper limit of camera feeds that can be monitored effectively is 16 feeds or less. The 16-feed condition was chosen based on this finding. The 8-feed condition was chosen to explore lower task demands, the 24-feed condition was chosen to investigate whether a higher number of feeds can be monitored, potentially increasing mental workload. Performance, subjective mental workload, and objective mental workload were recorded as dependent variables. Participants responded to traffic incidents with a keypress; specific incident types are described in the Experimental Task section. For each incident, the occurrence of a response (response vs. no response), the total response time (from incident start to response), and post-fixation response time (from the first fixation on the incident to response) were recorded as performance measures. The two response times were assumed to measure different performance components. Total response time may include visual search and other processes before first fixation, while post-fixation response time may reflect processes following initial visual detection. Blink rate and average pupil diameter were recorded as physiological measures for objective mental workload, and the NASA-TLX was used to assess subjective mental workload.

With an increase in task demands, attentional resource theory predicts subjective and objective mental workload to increase. At the same time, performance is predicted to decrease if demands exceed operator capacity. Consequently, subjective and objective mental workload were expected to increase as the number of CCTV feeds increased, as reflected by increasing NASA-TLX scores, decreasing blink rates and increasing pupil diameters. Performance was expected to decrease with an increase in monitored CCTV feeds, as reflected by increasing total and post-fixation response times and decreasing response counts. These effects, however, were only expected to occur if the monitored feeds provided sufficient information to maintain an adequate mental model as well as sufficient situational awareness of the situation. If the number of monitored feeds was too low, performance was also expected to decrease while mental workload increased, because of the added effort needed to compensate for missing information.

Methods

Participants

Participants were Dutch traffic control room operators recruited from several municipalities in the Netherlands. All were trained, with unimpaired or corrected vision. Sixteen participants were recruited, 15 men and one woman (Table 1). Participants had an average of M = 5.63 years (SD = 5.13 years) of experience as control room operators, with experience ranging from 0.5 to 15 years.

Table 1

Participant Age Distribution

Age group	*n*	%
Younger than 20	0	0.00
20-29	2	12.50
30-39	3	18.75
40-49	5	31.25
50-59	4	25.00
60 or older	2	6.70

Note. Age was grouped into predefined categories to maintain participant anonymity. Percentages show the proportion of participants from each age group compared to the total sample.

The study was approved by the Humanities & Social Sciences (HSS) Ethic Committee of the University of Twente on November 12th, 2024. A power analysis was conducted using G*Power version 3.1.9.7 (Faul et al., 2009), indicating the sample size needed to achieve 80% power for detecting a medium effect (Cohen, 1992) with a significance level of $\alpha = .05$ to be $N = 28$, while detecting a large effect ($f = 0.4$) would need $N = 12$ participants. This study was powered to reliably detect large effects, with smaller effects being harder to identify.

Materials

Experimental Task

For the experiment, participants were instructed to monitor 8, 16, or 24 CCTV feeds presented on a screen and press a key as quickly as possible once they noticed a traffic incident. The footage continued playing regardless of keypresses. Participants were informed that incidents could include vehicles stopping on the road or in the tunnel, traffic crashes, unauthorized users such as bikes or pedestrians entering the road or tunnel, and wrong-way drivers.

Measures

Subjective mental workload was measured using the general mental workload scores provided by the NASA-TLX. The scores ranged from 0 (*low mental workload*) to 100 (*high mental workload*). For this, the instructions provided by Hart (1986) were followed.

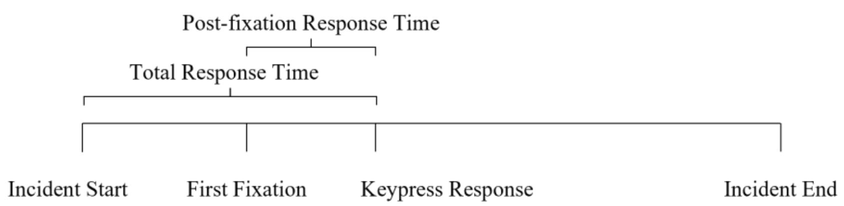
Average pupil diameter was calculated for the duration of the monitoring task. The raw pupil diameter values were recorded at a sampling rate of 50 Hz for each eye. They were averaged across both eyes to improve reliability. If a reading was missing for one eye at a given timestamp, the value measured from the other eye was used instead. Data points were only included in the calculation of average pupil diameter if at least one eye had a valid reading.

Blink rate, which refers to the number of eye blinks per minute, was measured for the duration of the monitoring task. Pupil diameter data was used to indirectly infer blinks (defined as the period in which both eyes are closed) using clusters of missing data for both eyes, as is common practice with some eye trackers (Nyström et al., 2024). Closed eyes resulted in missing pupil diameter data, for which a minimum duration of 60 milliseconds and a maximum duration of 400 milliseconds were set as criteria for blink detection. This was done following a combination of the method in Demiral et al. (2022) and a recommendation received in an email exchange with Tobii.

Two response times were calculated. Total response time was defined as the time from start of the time of interest (TOI) to the keypress (Figure 2). The TOI began at the first visible sign of an incident and ended at the end of the incident, which was 60 seconds later in most cases. The second measure, post-fixation response time, was defined as the time from a participant's first fixation on an incident to their keypress (Figure 2). All camera feeds showing an incident were defined as areas of interest (AOI) in their entirety, regardless of how much area of the feed showed the incident. The time to first fixation within the incident AOI was recorded for each incident. The post-fixation response time was calculated as the difference between the keypress timestamp and the first fixation timestamp.

Figure 2

Timeline of an Incident-TOI



The two measures served different purposes. Total response time was included as a more practical and comprehensive measure of performance in control rooms and was assumed to encompass both visual search and decision-making processes. The assumption was that the first fixation may reflect an initial response to the incident, followed by the keypress as a second, more deliberate response. By measuring from the beginning of an incident, it may provide a more complete representation of performance. Total response time is likely influenced by the number of monitored

CCTV feeds: as the number of feeds increases, the chances of immediately fixating the correct feed decrease, increasing the time until first fixation. In contrast, post-fixation response time was intended to reflect cognitive processes occurring after the visual detection of an incident. The measure was included to focus on the effects of task demands and the resulting potential increase in mental workload on decision-making processes. Under higher task demands, when monitoring more feeds for example, attentional resources are strained, as they are divided between feeds (Kahneman, 1973). This might leave fewer resources available for decision-making processes, which could lead to slower responses. Because this measure begins at the point of first fixation, it is not directly affected by differences in the time it takes to visually locate incidents.

The response count was manually calculated by counting the number of incidents that participants responded to. Each participant completed three 25-minute-long trials using three different videos, with three incidents shown per trial. The 25-minute duration was chosen due to time constraints and limited availability of real traffic footage. A response was recorded when a participant pressed the spacebar during the TOI defined for the incident.

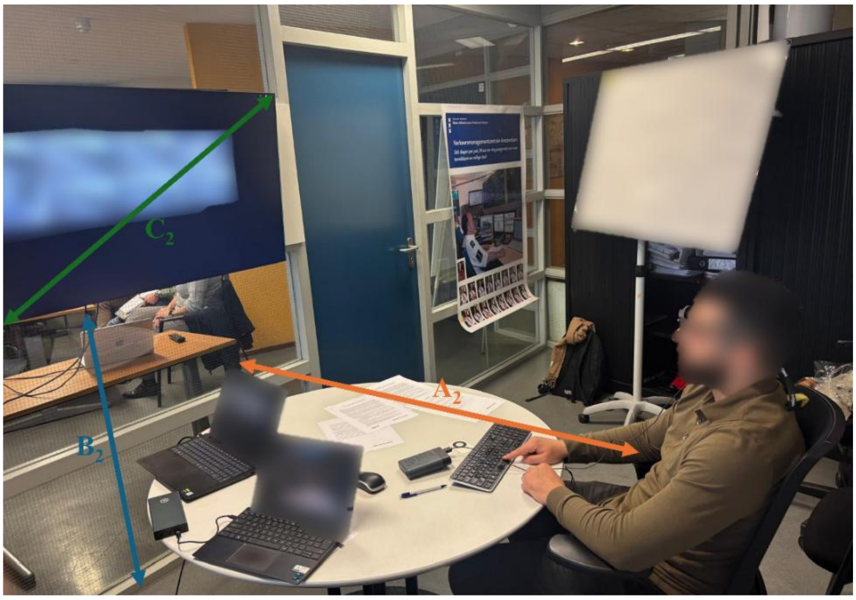
Subjective sleepiness was measured using self-reported sleepiness scores provided after each trial by the Karolinska Sleepiness Scale (KSS), a self-report measure (Åkerstedt & Gillberg, 1990). Scores range from 1 (*Extremely alert*) to 9 (*Extremely sleepy – fighting sleep*) (Figure A1).

Equipment and Setup

The experiment was conducted at four different locations to accommodate the participating municipalities and participants. To accomplish this, a mobile setup was used (Appendix D; Figure 3).

Figure 3

Example of an Experimental Setup



Note. This photo was taken at the second experiment location. Identifying details were blurred for privacy reasons.

General material. Eye-tracking data were recorded using Tobii Pro Glasses 2 (Tobii Pro, 2024a), a head-mounted eye tracker. The eye tracker was connected wirelessly to a Dell Vostro 3530 laptop (Intel Core i5-1335U, 8 GB RAM, Windows 11) for all participants at the first experiment location, and for three participants at the second location. The laptop was connected using a data cable for the remaining participants. It used the Tobii Pro Glasses Controller software (Tobii, 2019) to record eye-tracking data and control the calibration and recording processes. To play the traffic videos and run online questionnaires, a Dell Inspiron 7590 laptop (Intel Core I5-9300H, 16 GB RAM, Windows 11) was used. The laptops were not directly connected to each other but used at the same time during the experiment. A Dell KM714 wireless keyboard was connected to the Dell Inspiron 7590 laptop to capture participant key presses. Questionnaires were created and filled out online using the Alchemer Survey Software (Alchemer, 2024). Eye-tracking data was analyzed and exported using Tobii Pro Lab (Tobii Pro, 2024b). Traffic footage was cut and edited into the traffic videos using DaVinci Resolve (Blackmagic Desgin, 2024).

One participant information sheet (Figure B1) and one participant informed consent form (Figure B2) were created based on the informed consent form template for research with human participants (University of Twente, 2022). To make sure instructions were consistently given the same way, a paper copy of task instructions (Figure C1) was prepared. Two separate online questionnaires were created. Questionnaire A1 (Figure A1) included the NASA-TLX scale. The questionnaire also contained a question asking participants whether they had encountered this exact traffic situation or incident during their work as operators before, which was used as an exclusion criterion. The questionnaire was filled out after each trial. Questionnaire A2 (Figure A2) included the NASA-TLX pairwise comparison procedure. Additionally, in this questionnaire, participants were asked about their age, work experience, and whether they had any corrected or uncorrected vision impairments. All questionnaires, information forms, consent forms, and instructions were presented in Dutch. This was accomplished with the help of a Dutch native speaker at vhp human performance.

Traffic Incident Videos. For the experimental task, participants watched tunnel footage showing regular traffic, and nine traffic incidents. Footage from two separate tunnels (Tunnels 1 and 2) recorded by 24 different, approximately equally distanced traffic cameras was provided by two Dutch municipalities. Footage for each tunnel showed two separate tunnel tubes, leading traffic in opposite directions. The raw footage was split into three 25-minute-long separate videos (Videos A, B, and C) showing three traffic incidents each, combined with uneventful traffic footage. On average, an incident occurred approximately every eight minutes. This is a higher frequency than under real-world monitoring conditions, where incidents are usually separated by several hours of uneventful monitoring. Operators generally noted that it is impossible to estimate how often traffic incidents occur in practice. At least five minutes of uneventful footage was shown before and between incidents. One incident, shown in Video A, was recorded in Tunnel 2. The remaining eight incidents and uneventful footage were recorded in Tunnel 1. Each video was edited to match the three experimental conditions with eight feeds in the first condition, 16 in the second and 24 in the third condition. Each tube was shown by half of the feeds.

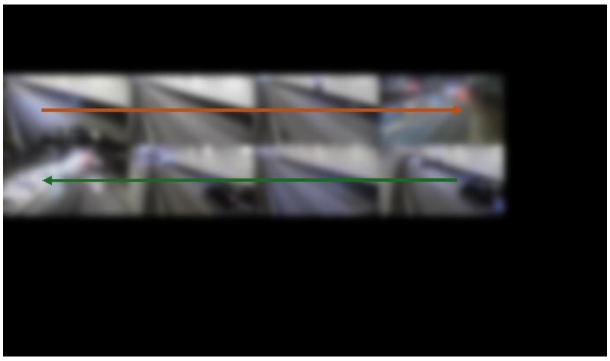
The arrangement of the video feeds aimed to create a layout of images resembling realistic conditions as closely as possible. This was done while accommodating the limitations of using one single screen instead of a video wall consisting of multiple comparable screens. Visual angles for screens and feeds in this experiment could not be matched to the corresponding visual angles in control rooms, as they vary considerably between different control rooms. The vertical viewing angles for each screen (Table D1) were in the range recommended by the Nederlands Normalisatie Instituut [NEN] (1997). Horizontal viewing angles exceeded the recommended range but were smaller than horizontal viewing angles observed in the control rooms visited for this study. Feeds from the same tube were grouped together in rows where possible and presented in the order of traffic flow. The feed layout (the way feeds were organized into rows on the screen) differed between the three conditions (Figure 4, Screenshots A, B, and C). Within each condition, the feed layouts of Tunnels 1 and 2 were the same, except in the 24-feed condition. In that case, a different layout was used for Tunnel 2 to fit all feeds onto the screen because of the differences in aspect ratio between feeds from Tunnels 1 and 2

(Figure 4, Screenshot D). Of the nine unique incidents shown during the experiment, only one was recorded at Tunnel 2. The rest were from Tunnel 1.

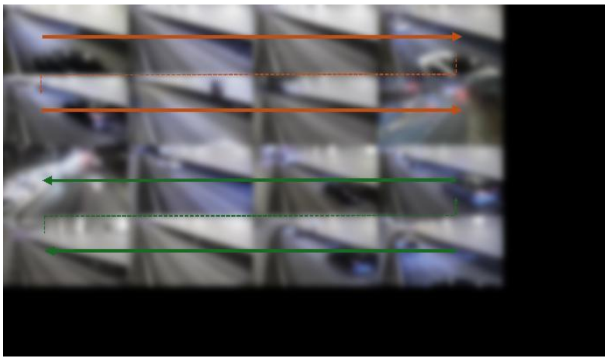
Figure 4

Feed Layouts

A



B



C

D

Note. Screenshots A-C show Tunnel 1 feed layouts for the 8, 16, and 24-feed conditions, respectively. Screenshot D shows the layout for Tunnel 2 in the 24-feed condition. Tunnel 2 was included in all three conditions, but the layouts for Tunnel 2 in the 8- and 16-feed conditions are not shown in this figure. These layouts had the same structure as those of Tunnel 1 (Screenshots A and B), the only difference was the feed aspect ratio. Arrows indicate traffic flow direction. Two colors represent the two tubes. The layouts shown in this thesis were blurred for privacy reasons.

Procedure

Participants signed consent forms and then received instructions for the monitoring task. Next, they were briefed on the use of the Tobii Pro Glasses 2, which they put on. The fit was verified by checking the eye images available in the Tobii Pro Glasses Controller software. The glasses were calibrated in the controller software. The calibration was then manually validated by instructing the participant to look at specific feeds and verifying the controller live view. For the task, participants were asked to keep one finger close to the spacebar. In the next step, the eye-tracking recording was started, and the first traffic footage video was played. Once the video had finished playing, participants took off the eye tracker and filled out Questionnaire A1, which included NASA-TLX and

KSS. These steps were repeated for each of the three trials, each trial corresponding to one of the three experimental conditions. Each trial consisted of 25 minutes of monitoring footage, during which three incidents were shown, followed by around five minutes to complete the questionnaires. After the completion of the third trial, participants filled out Questionnaire A2, which includes the NASA-TLX pairwise comparison procedure. After the conclusion of the experiment, participants were given the opportunity to ask final questions.

To account for order effects, counterbalancing was used to assign participants to different orders of the experimental conditions. In which order which videos (A, B, and C) were watched was also counterbalanced to avoid any biases. Each of the 16 participants was assigned to a unique combination of experimental condition order and video sequence.

Design

The study was a one-factor within-subject design. The independent variable was the number of CCTV feeds. The dependent variables were operator performance, subjective and objective mental workload. Subjective mental workload was assessed using the NASA-TLX. Objective mental workload was assessed using the physiological measures of blink rate and pupil diameter. Operator performance was assessed using total and post-fixation response times to incidents, and incident response counts.

Analysis

Exclusion Criteria and Missing Data

Outliers were detected using the interquartile range (IQR) method, identifying any outliers one and a half times the interquartile range above Q3, or one and a half times the interquartile range below Q1. Outliers were excluded unless they were inside plausible physiological limits and task observations. For example, pupil diameter values within the range found to be physiologically possible in past research (Spector, 1990) were not excluded. Eye-tracking recordings were excluded from further analysis of blink rate and pupil diameter if less than 80% of recorded gaze samples were valid. A gaze sample was valid when data for at least one eye was available for that time point. Whether recordings not meeting this criterion were still useful for the post-fixation response time analysis was decided using manual verification of the eye-tracking recordings, specifically focusing on incident TOI. Any response metric recorded from an incident which a participant had seen before was excluded from the analysis. Some responses were not recorded by the Tobii Pro Glasses Controller software because of technical issues. In these cases, the moment of response was estimated by logging the moment of the onset of the sound produced by pressing the spacebar, which was recorded by the eye tracker.

Analytic Strategy

Before any analysis, the assumptions for each model were tested. For linear and linear mixed effects models, the assumptions of normally distributed residuals (and random effects) were visually checked using quantile-quantile (Q-Q) plots and tested using the Shapiro-Wilk test. To evaluate linearity and independence, residuals vs. predicted values plots were created and checked. The assumption of homoscedasticity was verified by inspecting a boxplot of residuals by CCTV-feed condition. For the generalized linear mixed effects model, the normality of random effects was tested using the same method as for the linear mixed effects models.

Linear mixed effects models. For the analysis of blink rate, pupil diameter and NASA-TLX scores, linear mixed effects models were created. The number of CCTV feeds (with three levels: 8 feeds, 16 feeds, and 24 feeds) and subjective sleepiness scores were included as fixed effects variables. Participants were included as random effects. The number of CCTV feeds was included as a fixed effect to evaluate its impact on each outcome measure, while sleepiness was included as a fixed effect to control for its potential effect on the outcome measure. Participants were included as a random effect to consider individual variability. The model for NASA-TLX scores used ranked values instead of raw scores, as the NASA-TLX uses an ordinal scale, violating linear model assumptions. The following models were used in the analysis:

$$Outcome\ Measure_{ij} = \beta_0 + \beta_1\ (CCTVfeeds_i) + \beta_2\ (Sleepiness_{ij}) + \gamma_{0i} + \epsilon_{ij} \qquad (1)$$

In this model, Outcome Measure_{ij} stands for Blink Rate_{ij}, Pupil Diameter_{ij}, or Ranked NASA-TLX Score_{ij}, depending on the outcome measure used in each model. The components of the model are defined as follows:

- Outcome Measure_{ij} is the outcome measure of participant i in CCTV-feed condition j
- CCTV feeds_i is the number of CCTV feeds monitored by participant i
- Sleepiness_{ij} is the subjective sleepiness score of participants i after experiencing CCTV-feed condition j
- β_0 is the intercept
- β_1 is the fixed effect of the number of CCTV feeds on the outcome measure
- β_2 is the fixed effect of sleepiness on the outcome measure
- γ_{0i} is the random effect of each individual
- ϵ_{ij} is the residual error

Linear models. For the analysis of total and post-fixation response times, linear mixed models were created first, which resulted in a singular fit. This indicated that random effect variance was detected to be zero, with the participant random effects not adding to the models. To address the issue and avoid overfitting, the random effect was removed from the models, as recommended by Bates et al. (2015). After the removal, linear models were used to examine the fixed effects of the number of CCTV feeds and sleepiness. Due to skewed residuals, post-fixation and total response times were log-transformed to improve model fit. Additionally, to correct for recording delays that resulted in negative response times, a constant (one millisecond larger than the smallest negative response time value) was added to all response times before the transformation. Both response times were modeled with CCTV feeds (three levels: 8, 16, 24) and sleepiness scores as fixed effects. The models can be written as:

$$\log\left(Response\ Time_{ij}\right) = \beta_0 + \beta_1\,(CCTVfeeds_i) + \beta_2\,\left(Sleepiness_{ij}\right) + \epsilon_{ij} \qquad (2)$$

In this model, Response $Time_{ij}$ stands for Post-fixation Response $Time_{ij}$ or Total Response $Time_{ij}$, depending on the measure used in each model. The components of this model are defined as mentioned above.

Generalized linear mixed effects model. For the analysis of participant responses, a generalized linear mixed effects model with a binomial distribution and logit link function was created. It was modeled with the number of CCTV feeds (with three levels: 8, 16, and 24 feeds) and subjective sleepiness scores as fixed effects variables, and participants as random effects. The model was used to accurately represent the dichotomous nature of participant responses, and can be written as:

$$logit\left(P\left(Response_{ij} = 1\right)\right) = \beta_0 + \beta_1\,(CCTVfeeds_i) + \beta_2\,\left(Sleepiness_{ij}\right) + \gamma_{0i} \qquad (3)$$

The components of this model are defined as mentioned above. In this model specifically, $logit\left(P\left(Response_{ij} = 1\right)\right)$ is the log odds of participant i responding to an incident in CCTV-feed condition j.

For all models, experiment location was first included as a covariate to check for differences between experiment locations. There were no significant effects, and it was removed from the final model. Since a total of six different tests were performed to test mental workload and performance, a Bonferroni correction was applied to correct for the inflation of the alpha error, resulting in a significance level of $\alpha = 0.0083$.

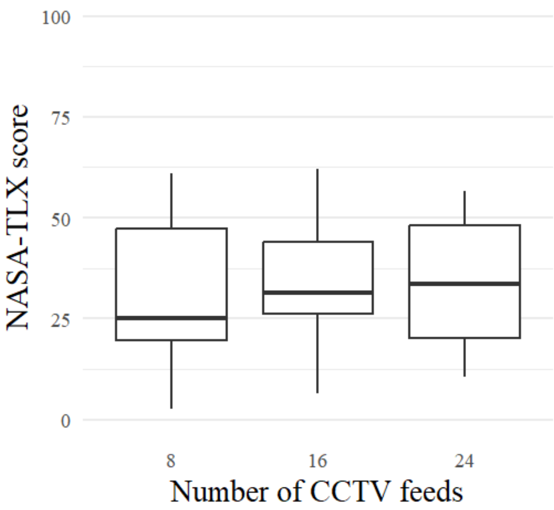
Results

NASA-TLX

To examine if NASA-TLX mental workload scores differed between CCTV-feed conditions, a linear mixed model was used. It included CCTV feeds (8 feeds, 16 feeds, 24 feeds) as a fixed categorical effect, sleepiness score as a fixed continuous effect, and random effects for participants. Type III Wald F tests with Kenward-Roger degrees of freedom showed that the null hypothesis could not be rejected, $F(2, 30) = 1.12$, $p = .338$, $\eta_p^2 = 0.07$, indicating that there was no significant effect of the number of CCTV feeds on subjective mental workload. General mental workload scores (Figure 5) had a median of *Mdn* = 25.17 (IQR = 19.58-47.17) in the 8-feed condition (*n* = 16), *Mdn* = 31.33 (IQR = 26.08-44.00) in the 16-feed condition (*n* = 16), and *Mdn* = 33.67 (IQR = 20.25-48.00) in the 24-feed condition (*n* = 16).

Figure 5

NASA-TLX Scores for Each of the CCTV-Feed Conditions



Note. The box represents the interquartile range (IQR), the median NASA Task Load Index (NASA-TLX) score is represented by a horizontal line inside the box for each condition. The whiskers reach the lowest and highest value inside one and a half times the IQR below Q1, or one and a half times the IQR above Q3.

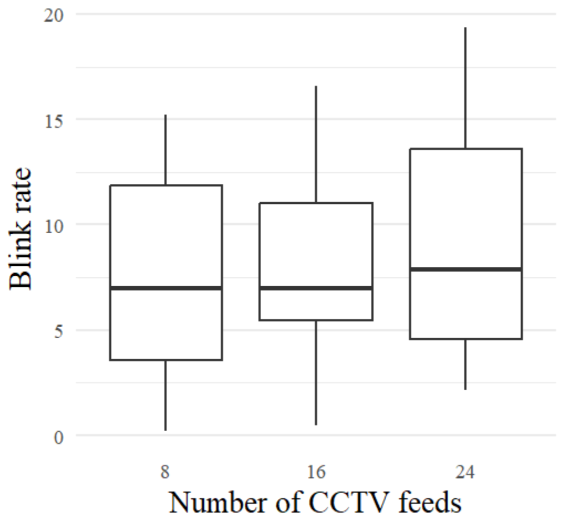
Blink Rate

The eye-tracking recording for one trial in the 8-feed condition was lost because of a recording error. Of the remaining 47 recordings, three recordings with gaze sample rates lower than 80% were excluded. One outlier was detected with the IQR method. The value was removed from the analysis as maintaining a blink rate of 33.47 blinks per minute over a period of over 25 minutes was deemed unlikely under normal conditions, considering the same participant’s blink rates in other conditions were 15.19 and 14.36 blinks per minute. While contact lens use can increase blink frequency, the large discrepancy between this value and the other recordings of the same participant suggests that the outlier likely occurred because of a measurement error.

To examine if blink rate differed between CCTV-feed conditions, a linear mixed model was used. It included CCTV feeds (8 feeds, 16 feeds, 24 feeds) as a fixed categorical effect, sleepiness score as a fixed continuous effect, and random effects for participants. Type III Wald F tests with Kenward-Roger degrees of freedom showed that the null hypothesis could not be rejected, $F(2, 25) = 1.30$, $p = .291$, $\eta_p^2 = 0.09$, indicating that there was no significant effect of the number of CCTV feeds on blink rate. The median blink rates (Figure 6) were $Mdn = 6.98$ (IQR = 3.58-11.84) blinks per minute in the 8-feed condition ($n = 15$), $Mdn = 6.98$ (IQR = 5.43-11.02) blinks per minute in the 16-feed condition ($n = 12$), and $Mdn = 7.87$ (IQR = 4.59-13.62) blinks per minute in the 24-feed condition ($n = 16$).

Figure 6

Blink Rates Across CCTV-feed Conditions



Note. The box represents the interquartile range (IQR), the median blink rate is represented by a horizontal line inside the box for each condition. The whiskers reach the lowest and highest value inside one and a half times the IQR below Q1, or one and a half times the IQR above Q3. One outlier was removed.

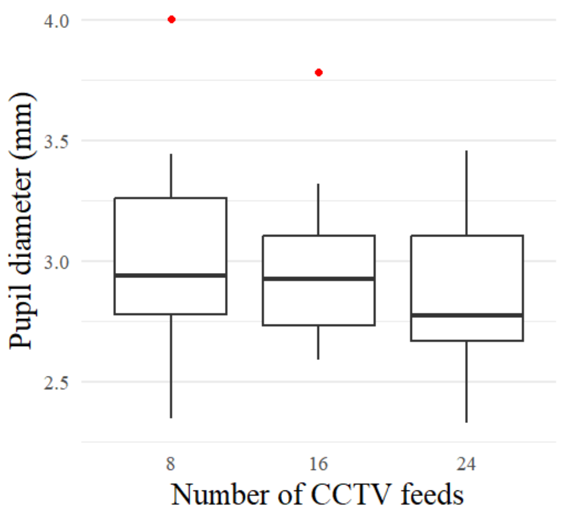
Pupil Diameter

The recordings excluded from the blink rate analysis due to low gaze sample rates or recording errors were also excluded from the pupil diameter analysis. Two outliers were detected with the IQR method. The values were not removed from the analysis as average pupil diameter values of 4.01 mm and 3.78 mm over 25 minutes were physiologically possible (Spector, 1990) and were measured for the same participant. A Q-Q plot showed possible minor deviations of residuals from the normal distribution in the 24-feed condition. A boxplot of residuals by CCTV-feed condition was visually checked and showed a slightly lower variation in residuals in the 24-feed condition, indicating a minor violation of homoscedasticity. Because of the non-severe nature of the violations, analysis was continued using the present model.

To examine if pupil diameter differed between CCTV-feed conditions, a linear mixed model was used. It included CCTV feeds (8 feeds, 16 feeds, 24 feeds) as a fixed categorical effect, sleepiness score as a fixed continuous effect, and random effects for participants. Type III Wald F tests with Kenward-Roger degrees of freedom showed that the hypothesis is accepted, $F(2, 25) = 9.19$, $p < .001$, $\eta_p^2 = 0.42$, indicating that there was a significant effect of the number of CCTV feeds on pupil diameter. Post-hoc pairwise comparisons with a family-wise Type I error rate of 0.05 and Bonferroni correction were used and showed no difference between pupil diameters in the 8-feed condition and the 16-feed condition ($p = .139$), or the 16-feed condition and the 24-feed condition ($p = .265$). They did show a difference between the 8-feed condition and the 24-feed condition ($p < .001$), indicating that participants in the 24-feed condition had smaller average pupil diameters over the duration of the monitoring task than in the 8-feed condition. The median average pupil diameters (Figure 7) were *Mdn* = 2.94 mm (IQR = 2.78-3.26 mm) in the 8-feed condition ($n = 15$), *Mdn* = 2.93 mm (IQR = 2.73–3.11 mm) in the 16-feed condition ($n = 13$), and *Mdn* = 2.77 mm (IQR = 2.67–3.10 mm) in the 24-feed condition ($n = 16$).

Figure 7

Average Pupil Diameter (mm) Across CCTV-feed Conditions



Note. The box represents the interquartile range (IQR), the median average pupil diameter is represented by a horizontal line inside the box for each condition. The whiskers reach the lowest and highest value inside one and a half times the IQR below Q1, or one and a half times the IQR above Q3. Outliers outside this range are shown as red dots.

Response Count

One incident was excluded because one-third of participants hesitated before responding, and two fixated on the incident without reacting, raising questions about the clarity of the incident. A check of the normality assumption for random effects using a Q-Q plot and the Shapiro-Wilk test indicated a possible violation of the normality assumption ($W = 0.851$, $p = .014$). The Q-Q plot did not show any violation of normality. Given that the model is robust to minor violations of the normality assumption for random effects, the model was still used.

To examine if the number of responses to incidents differed between CCTV-feed conditions, a generalized linear mixed effects model was used. It included CCTV feeds (8 feeds, 16 feeds, 24 feeds) as a fixed categorical effect, sleepiness score as a fixed continuous effect, and random effects for participants. A likelihood ratio test ($\chi^2(1) = 0.40$, $p = .820$) revealed no significant effect of the number of CCTV feeds on the model, suggesting that the number of CCTV feeds did not meaningfully contribute to explaining response count. Participants generally responded to incidents at a high rate, with a considerable percentage of participants not missing a single incident in a given trial (Table 2).

Table 2

Response Percentages

CCTV Feeds	Response Rate (%)	Complete Detection (%)	*n*
8	82.5	66.7	40
16	86.0	62.5	43
24	81.0	50.0	42

Note. Response rate represents the percentage of incidents that participants responded to in each condition. Complete detection represents the percentage of participants per condition that did not miss a single incident in that condition. An exploratory mixed-effects model predicting complete detection with the number of feeds and sleepiness as fixed effects did not show significant results.

Response Time

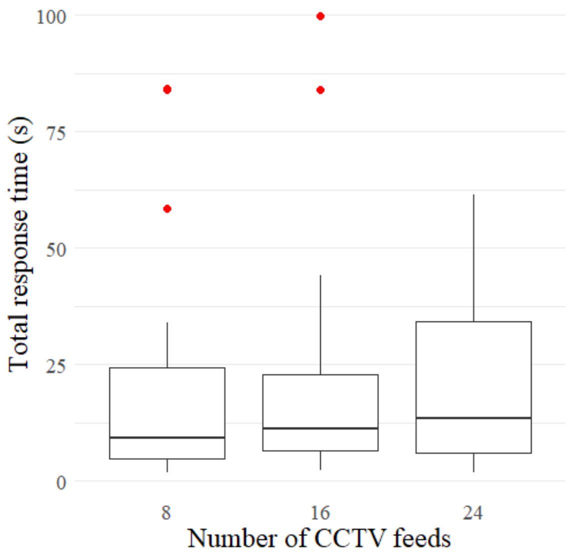
Total Response Time

The same incident as in the response count analysis was removed. Three responses were excluded based on participants reporting having previously seen the incident, resulting in 99 remaining response times. Six outliers were detected with the IQR method but were not removed from the analysis, as they likely represented naturally occurring variations in response behavior.

To examine if total response time differed between CCTV-feed conditions, a linear model was used. It included CCTV feeds (8 feeds, 16 feeds, 24 feeds) as a fixed categorical effect and sleepiness score as a fixed continuous effect. Type III Wald F tests showed that the null hypothesis could not be rejected, $F(2, 95) = 0.28$, $p = .757$, $\eta_p^2 = 0.01$, indicating that there was no significant effect of the number of CCTV feeds on response times. The median response times (Figure 8) were *Mdn* = 9.25 s (IQR = 4.77–24.24 s) in the 8-feed condition ($n = 31$), *Mdn* = 11.10 s (IQR = 6.40–22.84 s) in the 16-feed condition ($n = 35$), and *Mdn* = 13.46 s (IQR = 5.86–34.28 s) in the 24-feed condition ($n = 33$).

Figure 8

Total Response Time to Incidents Across CCTV-feed Conditions



Note. The box represents the interquartile range (IQR), the median total response time is represented by a horizontal line inside the box for each condition. The whiskers reach the lowest and highest value inside one and a half times the IQR below Q1, or one and a half times the IQR above Q3. Outliers outside this range are shown as red dots.

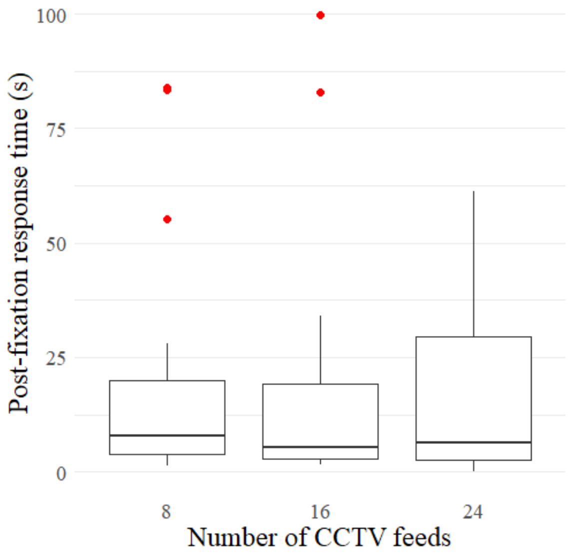
Post-fixation Response Time

The same incident as in the response count analysis was removed. Six response times were excluded from the response time analysis due to insufficient recording quality during incident TOI. Another three responses were excluded based on participants reporting having previously seen the incident, resulting in 94 remaining response times. Six outliers were detected with the IQR method but were not removed from the analysis, as they likely represented naturally occurring variations in response behavior.

To examine if response time differed between CCTV-feed conditions, a linear model was used. It included CCTV feeds (8 feeds, 16 feeds, 24 feeds) as a fixed categorical effect and sleepiness score as a fixed continuous effect. Type III Wald F tests showed that the null hypothesis could not be rejected, $F(2, 90) = 0.24$, $p = .790$, $\eta_p^2 = 0.00$, indicating that there was no significant effect of the number of CCTV feeds on post-fixation response times. The median response times (Figure 9) were *Mdn* = 7.91 s (IQR = 3.75–19.93 s) in the 8-feed condition ($n = 31$), *Mdn* = 5.29 s (IQR = 2.76–19.14 s) in the 16-feed condition ($n = 31$), and *Mdn* = 6.40 s (IQR = 2.56–29.55 s) in the 24-feed condition ($n = 32$).

Figure 9

Post-fixation Response Time to Incidents Across CCTV-feed Conditions



Note. The box represents the interquartile range (IQR), the median post-fixation response time is represented by a horizontal line inside the box for each condition. The whiskers reach the lowest and highest value inside one and a half times the IQR below Q1, or one and a half times the IQR above Q3. Outliers outside this range are shown as red dots.

Sleepiness

Sleepiness was included as a covariate in all models but was not shown to significantly impact any of the outcome measures.

Discussion

A vigilance monitoring experiment was conducted with the number of CCTV feeds (8, 16 or 24) being varied within participants as the independent variable. Participants responded to traffic incidents with a keypress. Total and post-fixation response times, as well as response counts were recorded as performance measures. Blink rate and average pupil diameter were recorded as physiological measures for objective mental workload, and the NASA-TLX was used to assess subjective mental workload. An increase in mental workload, as well as a decrease in performance were expected with an increase in monitored CCTV feeds. These expected effects were not confirmed in the experiment, suggesting that operators were able to handle monitoring up to 24 feeds without noticeable increases in mental workload or performance decrements.

Interpretation

The number of monitored CCTV feeds did not influence subjective mental workload or performance, not supporting the hypotheses that an increase of the number of CCTV feeds increases subjective mental workload and decreases performance. However, a main effect of the number of monitored CCTV feeds on pupil diameter was found. Post-hoc pairwise comparisons between conditions found average pupil diameters over the duration of the monitoring task in the 8-feed condition to be larger than in the 24-feed condition, which indicates a direction of effect opposite of the hypothesis. Blink rate was not affected by feed count. This does not support the hypothesis that objective mental workload increases as the number of CCTV feeds increases.

The findings contrast with past research linking increasing task demands and with higher subjective mental workload (Warm et al., 1996), decreased performance (with studies mainly considering response count and total response time) (Helton & Russell, 2013; Helton & Warm, 2008), and lower blink rate (Biondi et al., 2023; Stern et al., 1994). The results also contradict findings that higher numbers of CCTV feeds lead to a decrease in performance (Gill et al., 2005), with a maximum number of 12-16 feeds being recommended (Wallace et al., 1997, as cited in Diffley & Wallace, 1998; Tickner & Poulton, 1973). In addition, the present research contradicts past findings of an effect of task demands on pupil diameter (van der Wel & van Steenbergen, 2018), as the effect observed in the present study was in the opposite direction.

An important explanation of these inconsistencies could be differences in the way task demands were manipulated in past studies and the present experiment. Past studies increased the number of feeds and scenes at the same time, while both factors are known to increase mental workload and decrease performance (Wallace et al., 1997, as cited in Diffley & Wallace, 1998; Pikaar, 2015; Tickner & Poulton, 1973; Wood, 2007). In addition to that, they mainly focused on security monitoring contexts and used complex CCTV footage showing footage from town centers or prisons. In contrast, the present experiment increased only the number of feeds, while keeping the number of scenes constant at two. The footage shown was also less complex, with objects in most feeds moving in one direction only. These differences may have led to smaller differences in task demands between conditions.

According to attentional resource theory, increasing the number of monitored feeds (and therefore task demands) results in higher mental workload and decreasing performance if task demands exceed operator capacity. For this to be true, however, a substantial increase in required attentional resources is assumed. In the present study, the difference in the number of feeds between conditions, especially considering that only two scenes were monitored, may not have been large enough to cause a strong difference in required attentional resources. This would explain the lack of observed differences in mental workload or performance. Consequently, past recommendations of a maximum of 16 monitored feeds (Velastin, 2003, as cited in Schreibers et al., 2012; Wallace et al., 1997, as cited in Schreibers et al., 2012; Wood, 2007) and findings of considerable differences in detection rates between 9 and 16 monitored feeds (Tickner & Poulton, 1973) are contradicted by the present study, as it ranged from 8 to 24 feeds. Because of the differences in task demands and the number of monitored scenes between different control room contexts, generalized recommendations for a maximum number of monitored feeds may not be applicable for all control rooms. Following up on this study, future studies could use larger differences in the number of feeds between conditions to create larger differences in task demands. Especially varying the number of monitored scenes could be of high relevance for tunnel traffic control rooms.

The participation of trained operators likely also contributes to explaining why increases in mental workload and decreases in performance with rising task demands were not observed, in contrast with past research on objective and subjective mental workload and performance which mainly involved untrained participants (Biondi et al., 2023; Wallace et al., 1997, as cited in Diffley & Wallace, 1998; Helton & Russell, 2013; Helton & Warm, 2008; Tickner & Poulton, 1973; van der Wel & van Steenbergen, 2018). Trained operators may use more automated cognitive processes during monitoring than untrained participants, reducing the attentional resources required for the task (Young et al., 2015). As a result, their mental workload is lower and their performance higher compared to untrained participants (Young et al., 2015), allowing them to handle changes in demands, that would have larger effects on the performance and mental workload of untrained participants.

Additionally, task demands may have been in an optimal mental workload and performance range (Figure 1), leaving operators with sufficient resources to complete the task. This view does not align with Warm et al. (2008), who argued that vigilance tasks require mental work and are stressful. However, the task being not challenging enough would explain the good performance across conditions (Table 2), as well as the lack of differences in performance between them. It would also explain the moderate NASA-TLX scores indicating neither overload nor underload. Other studies presenting incidents more frequently (although this does not reflect real-world monitoring conditions) support this view, with some experiments showing an incident around every 4 minutes (Tickner & Poulton, 1973), and some showing incidents at high rates between 5 and 40 incidents per minute (Warm et al., 1996), which increases mental workload. In addition to showing incidents more frequently, some studies that found higher mental workload used longer monitoring task durations of up to 60 minutes per condition (Tickner & Poulton, 1973; Warm et al., 1996). In future research, it may be promising to increase task demands by increasing monitoring duration and adding more incidents, feeds or scenes. A longer task duration could lead to operators experiencing sustained high mental workload over time, potentially revealing effects of overload that were not observed in this study because of its relatively short monitoring duration.

Furthermore, because operators were not instructed to follow any monitoring strategy, they may have prioritized some feeds over others. As a result, the number of frequently monitored feeds would have been reduced to few high-priority feeds, while low-priority feeds were infrequently scanned. This may have resulted in reduced required attentional resources in higher CCTV-conditions, and lower mental workload. Although this strategy does not reliably improve performance (and therefore cannot explain why no effects on performance were found), it is commonly used among operators (Hodgetts et al., 2018; Tickner & Poulton, 1973; Wickens et al., 2021). Tickner and Poulton (1973) found trained operators to be better at selecting priority feeds than untrained participants, offering one possible explanation to the results of the present study contrasting with past research. Overall, the self-directed use of different strategies by operators may have added variability to the results. Providing more specific instructions could be a viable alternative for future research.

Participants fixating incident footage but then moving on to monitor different feeds and responding later may have led to the lack of differences in post-fixation response times between conditions. This observation could point towards the phenomenon of inattentional blindness (Mack, 2003). Inattentional blindness refers to the phenomenon of failing to notice stimuli, even when they are directly in line of sight. Since post-fixation response times were measured from the moment of first fixation on an incident, this behavior may have inflated post-fixation response times overall and added variation, making differences in post-fixation response times between conditions harder to detect. Still, there was also no effect of the number of feeds on total response time, indicating that the absence of expected effects was not limited to gaze-based response times. Further studies investigating inattentional blindness with the help of eye tracking could deliver knowledge which could be used to mitigate the risk of missed incidents in the control rooms.

Moreover, averaging measures for objective mental workload may have masked differences. Task demands may have caused larger differences in mental workload between conditions at the beginning of the experiment. However, with the task progressing, the difference then may have shrunk after resources were depleted for all conditions. As a result, averaging could have led to smaller overall differences in pupil diameter or blink rate over the entirety of a trial. Other studies that reported a link between pupil diameter or blink rate and task demands used averages for shorter task lengths (Biondi et al., 2023; Stern et al., 1994; Veltman & Gaillard, 1996), possibly counteracting this issue. Analyzing objective measures for different parts of trials separately could be a viable alternative. Similar limitations are true for the measurement of subjective mental workload. In contrast to objective mental workload, which is measured over the entire duration of each trial, subjective mental workload is assessed once at the end of a trial. Although this is simple and nonintrusive, it results in a loss of information. Assessing subjective mental workload more frequently during trials (e.g. after incidents) using methods like instantaneous self-assessment (Tattersall & Foord, 1996) could help mitigate the loss of information.

Another limitation of the study is its reliance on one primary task as a performance measure of mental workload. For future research, adding a secondary task to the primary task could strengthen the measurement of mental workload, as it can offer insights into the remaining attentional capacities unused by the primary task (Young et al., 2015). It could also increase external validity, as operators rarely focus only on monitoring.

The mobile experimental setup allowed for the flexible and practical implementation of the experiment, but may also have introduced additional variability, especially in pupil diameter. With lighting impacting pupil diameter through the pupillary light reflex (Ellis, 1981; Pfleging et al., 2016), screen brightness may have influenced pupil diameters. As the number of feeds increased, the number of illuminated pixels also increased because of constant feed size regardless of feed count (Figure 4). This may have resulted in smaller pupil diameters in conditions with more feeds. This may explain why average pupil diameter was significantly smaller in the 24-feed condition than in the 8-feed condition (Figure 7). Consequently, the expected increase of pupil diameters in higher workload conditions may have been counteracted by this effect. The present experiment hints at challenges concerning the use of pupillometry in real-world scenarios to assess mental workload. Although pupil diameter has been shown to be related to mental workload under more controlled lighting conditions (van der Wel & van Steenbergen, 2018), and is even recommended by manufacturers for assessing mental workload in applied settings (Tobii, 2024), the results of this study do not support this recommendation.

An additional limitation of using a mobile setup was the use of a head-mounted eye tracker, which in combination with changing environments, resulted in sometimes suboptimal tracking accuracies. Because of this, no distinction was made between looking at an incident versus looking at the feed that contained the incident. In some cases, operators may have looked at the feed containing the incident, while not fixating the incident directly, which would have been counted as still looking at the incident. Careful calibration and the use of more advanced eye-tracking equipment could resolve these inaccuracies. Furthermore, because of the practical constraints of using a mobile setup, all feeds were displayed on a single screen (Figure 3), although operators typically monitor entire video walls comprised of multiple screens. For future research, traffic footage could be presented on larger video walls, increasing external validity by making the experiment more similar to real-world conditions.

Conclusion

The present study did not find any significant effect of the number of monitored CCTV feeds on mental workload or operator performance when traffic was presented across two distinct scenes (tunnel tubes). Operators were able to monitor traffic within the range of 8, 16, and 24 feeds across two scenes while maintaining high performance and experiencing no substantial increases in mental workload. Considering these results, past findings limiting the number of effectively monitored feeds to a maximum of 16 feeds per operator may not apply to all control room contexts, particularly tunnel traffic control rooms. Consistent with Pikaar (2015), this can be explained by not only the number of feeds but also feed content, specifically the number of distinct scenes, being an important factor influencing mental workload and performance. A majority of past research manipulated the number of feeds and distinct scenes simultaneously, while the present study isolated the effect of the number of feeds alone by keeping the number of distinct monitored scenes consistent. Additionally, operator experience and the automation of cognitive processes needed for monitoring may have contributed to operators being able to successfully monitor more feeds than expected. The use of trained and experienced operators, combined with the fact that only two scenes were monitored in all conditions, may have reduced attentional resource demands and minimized differences in demands between conditions. This could have prevented measurable differences in performance or mental workload.

While operators maintained good performance overall, most were not able to detect all incidents that occurred. The present findings support the importance of involving training and systems that reduce the risk of missing incidents and mitigate its consequences when it occurs. Especially the use of automated systems alerting the operator to potential incidents could be an adequate way of dealing with this risk. Increasing the integration of automated systems in control rooms could help offload vigilance tasks to machines (Warm et al., 1996), allowing operators to take on a more supervisory role rather than mainly focusing on monitoring.

References

Akca, M., & Küçükoğlu, M. T. (2020). Relationships between mental workload, burnout, and job performance: A research among academicians. In *Evaluating mental workload for improved workplace performance* (pp. 49-68). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-7998-1052-0.ch003

Åkerstedt, T., & Gillberg, M. (1990). Subjective and Objective Sleepiness in the Active Individual. *International Journal of Neuroscience*, *52*(1-2), 29-37. https://doi.org/10.3109/00207459008994241

Alchemer. (2024). *Alchemer Survey*. In (Version 2024.12.09.00) Alchemer. https://app.alchemer.eu/

Barbato, G., De Padova, V., Paolillo, A. R., Arpaia, L., Russo, E., & Ficca, G. (2007). Increased spontaneous eye blink rate following prolonged wakefulness. *Physiol Behav*, *90*(1), 151-154. https://doi.org/10.1016/j.physbeh.2006.09.023

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*. https://doi.org/https://doi.org/10.48550/arXiv.1506.04967

Biondi, F. N., Saberi, B., Graf, F., Cort, J., Pillai, P., & Balasingam, B. (2023). Distracted worker: Using pupil size and blink rate to detect cognitive load during manufacturing tasks. *Applied Ergonomics*, *106*, 103867. https://doi.org/https://doi.org/10.1016/j.apergo.2022.103867

Blackmagic Desgin. (2024). *DaVinci Resolve*. In (Version Version 19.1) [Computer software]. Blackmagic Design. https://www.blackmagicdesign.com/products/davinciresolve

Brookings. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3). https://doi.org/info:doi/10.1016/0301-0511(95)05167-8

Caffier, P. P., Erdmann, U., & Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, *89*(3), 319-325. https://doi.org/10.1007/s00421-003-0807-5

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics*, *74*, 221-232. https://doi.org/https://doi.org/10.1016/j.apergo.2018.08.028

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159. https://doi.org/10.1037//0033-2909.112.1.155

Davies, D. R., & Parasuraman, R. (1982). *The psychology of vigilance*. Academic Press. https://cir.nii.ac.jp/crid/1130282268955339264

Demiral, Ş. B., Manza, P., Biesecker, E., Wiers, C., Shokri-Kojori, E., McPherson, K., Dennis, E., Johnson, A., Tomasi, D., Wang, G.-J., & Volkow, N. D. (2022). Striatal D1 and D2 receptor availability are selectively associated with eye-blink rates after methylphenidate treatment. *Communications Biology*, *5*(1), 1015. https://doi.org/10.1038/s42003-022-03979-5

Diffley, C., & Wallace, E. (1998). *CCTV Control Room Ergonomics* (CCTV: Making it work, Issue 14/98). https://comfortzone-cctv.co.uk/wp-content/uploads/2019/04/Home-Office-CCTV-Control-Room-Ergonomics-14-98.pdf

Ellis, C. J. (1981). The pupillary light reflex in normal subjects. *The British Journal of Ophthalmology*, *65*(11), 754-759. https://doi.org/10.1136/bjo.65.11.754

Endsley, M. R., & Garland, D. J. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation awareness analysis and measurement*, *1*(1), 3-21. https://doi.org/10.1201/b12461

Esterman, M., & Rothlein, D. (2019). Models of sustained attention. *Current Opinion in Psychology*, *29*, 174-180. https://doi.org/https://doi.org/10.1016/j.copsyc.2019.03.005

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Gill, M., Spriggs, A., Allen, J., Hemming, M., Jessiman, P., Kara, D., Kilworth, J., Little, R., & Swain, D. (2005). Control room operation: findings from control room observations. *Home office online report*, *14*(05). https://tinyurl.com/rzkrmdxn

Halverson, T., Estepp, J., Christensen, J., & Monnin, J. (2012). Classifying Workload with Eye Movements in a Complex Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 168-172. https://doi.org/10.1177/1071181312561012

Hancock, P. A. (1989). A Dynamic Model of Stress and Sustained Attention. *Human Factors*, *31*(5), 519-537. https://doi.org/10.1177/001872088903100503

Hart, S. G. (1986). NASA Task Load Index (TLX): Paper and Pencil Package-Volume 1.0. https://humansystems.arc.nasa.gov/groups/tlx/downloads/TLX_pappen_manual.pdf

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Advances in Psychology* (Vol. 52, pp. 139-183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Helton, W. S., & Russell, P. N. (2013). Visuospatial and verbal working memory load: Effects on visuospatial vigilance. *Experimental Brain Research*, *224*(3), 429-436. https://doi.org/10.1007/s00221-012-3322-2

Helton, W. S., & Warm, J. S. (2008). Signal salience and the mindlessness theory of vigilance. *Acta Psychologica*, *129*(1), 18-25. https://doi.org/10.1016/j.actpsy.2008.04.002

Hodgetts, H., Chamberland, C., Latulippe-Thériault, J.-D., Vachon, F., & Tremblay, S. (2018). Priority or Parity? Scanning Strategies and Detection Performance of Novice Operators in Urban Surveillance. *Human Factors and Ergonomics Society Annual Meeting Proceedings*, *62*, 1113-1117. https://doi.org/10.1177/1541931218621255

Jalali, M., Esmaeili, R., Habibi, E., Alizadeh, M., & Karimi, A. (2023). Mental workload profile and its relationship with presenteeism, absenteeism and job performance among surgeons: The mediating role of occupational fatigue. *Heliyon*, *9*(9). https://doi.org/10.1016/j.heliyon.2023.e19258

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall. https://books.google.nl/books?id=tkvuAAAAMAAJ

Klonowicz, T. (1995). Mental Workload and Health: A Latent Threat. *International Journal of Occupational Safety and Ergonomics*, *1*(2), 130-135. https://doi.org/10.1080/10803548.1995.11076309

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451. https://doi.org/10.1037//0096-1523.21.3.451

MacDonald, W. (1999). Train controller interface design: factors influencing mental workload. 1999 International Conference on Human Interfaces in Control Rooms, Cockpits and Command Centres, Bath, UK. https://doi.org/10.1049/cp:19990159

Mack, A. (2003). Inattentional Blindness:Looking Without Seeing. *Current Directions in Psychological Science*, *12*(5), 180-184. https://doi.org/10.1111/1467-8721.01256

Mackworth, N. H. (1948). The Breakdown of Vigilance during Prolonged Visual Search. *Quarterly journal of experimental psychology*, *1*(1), 6-21. https://doi.org/10.1080/17470214808416738

Morad, Y., Barkana, Y., Zadok, D., Hartstein, M., Pras, E., & Bar-Dayan, Y. (2009). Ocular parameters as an objective tool for the assessment of truck drivers fatigue. *Accident; analysis and prevention*, *41*(4), 856-860. https://doi.org/10.1016/j.aap.2009.04.016

Moray, N. (1999). Mental models in theory and practice. In *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*. (pp. 223-258). The MIT Press. https://www.researchgate.net/publication/289336226_Mental_models_in_theory_and_practice

Nederlands Normalisatie Instituut [NEN]. (1997). Veiligheid van machines - Ergonomische eisen voor het ontwerpen van informatie- en bedieningsmiddelen - Deel 2: Informatiemiddelen. In: NEN.

Neil, D., Thomas, N., & Baker, B. (2007). Threat image projection in CCTV. 2007 41st Annual IEEE International Carnahan Conference on Security Technology, Ottawa. https://doi.org/10.1109/CCST.2007.4373500

Nyström, M., Andersson, R., Niehorster, D. C., Hessels, R. S., & Hooge, I. T. C. (2024). What is a blink? Classifying and characterizing blinks in eye openness signals. *Behavior Research Methods*, *56*(4), 3280-3299. https://doi.org/10.3758/s13428-023-02333-9

Pfleging, B., Fekety, D. K., Schmidt, A., & Kun, A. L. (2016). A model relating pupil diameter to mental workload and lighting conditions. Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, California. https://doi.org/10.1145/2858036.2858117

Pikaar, R. N. e. (2015). *Human Factors Guidelines for the Design of CCTV-Systems (Revision 1)*. E. H. F. Engineering.

Rostami, F., Babaei-Pouya, A., Teimori-Boghsani, G., Jahangirimehr, A., Mehri, Z., & Feiz-Arefi, M. (2021). Mental Workload and Job Satisfaction in Healthcare Workers: The Moderating Role of Job Control. *Frontiers in Public Health*, *9*. https://doi.org/10.3389/fpubh.2021.683388

Schreibers, K., Landman, R., & Pikaar, R. (2012). Human Factors of CCTV; Part 1. Technology and Literature review. *pp. S386–R1. CCT). Enschede: ErgoS Engineering and Ergonimics*.

Schreibers, K. B. J., & Bouchier, J. L. A. (2014). *CCTV - Case study traffic management highway tunnel*. 11th International Symposium on Human Factors in Organisational Design and Management & 46th Annual Nordic Ergonomics Society Conference, Copenhagen, Denmark. https://humansafetyalliance.com/wp-content/uploads/2020/11/2014_CCTV-Casestudy_ODAM2014.pdf

Spector, R. (1990). The Pupils. In H. H. Walker, WD; Hurst, JW (Ed.), *Clinical Methods: The History, Physical, and Laboratory Examinations* (3rd edition ed.). Butterworths. https://www.ncbi.nlm.nih.gov/books/NBK381/

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink Rate: A Possible Measure of Fatigue. *Human Factors*, *36*(2), 285-297. https://doi.org/10.1177/001872089403600209

Tattersall, A. J., & Foord, P. S. (1996). An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics*, *39*(5), 740-748. https://doi.org/10.1080/00140139608964495

Tickner, A., & Poulton, E. (1973). Monitoring up to 16 synthetic television pictures showing a great deal of movement. *Ergonomics*, *16*(4), 381-401. https://doi.org/10.1080/00140137308924529

Tobii. (2019). *Tobii Pro Glasses Controller*. In (Version 1.114.20033) [Computer software]. Tobii. https://connect.tobii.com/s/g2-downloads?language=en_US

Tobii. (2024). *Eye Tracking in Human Factors: Assessing Cognitive Abilities*. Tobii. https://www.tobii.com/resource-center/reports-and-papers/eye-tracking-in-human-factors#cta-section

Tobii Pro. (2024a). *Tobii Pro Glasses 2*. Tobii. Retrieved 24 September 2024 from https://www.tobii.com/products/discontinued/tobii-pro-glasses-2

Tobii Pro. (2024b). *Tobii Pro Lab*. In (Version 24.21.435) [Computer software]. Tobii. https://connect.tobii.com/s/lab-downloads?language=en_US

University of Twente. (2022). *Informed consent form template for research with human participants*. University of Twente. https://www.utwente.nl/en/bms/research/forms-and-downloads/informed-consent-form-template-for-research-with-human-participants-bms-2022.docx

van der Wel, P., & van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic Bulletin & Review*, *25*(6), 2005-2015. https://doi.org/10.3758/s13423-018-1432-y

Veltman, J. A., & Gaillard, A. W. K. (1996). Physiological indices of workload in a simulated flight task. *Biological Psychology*, *42*(3), 323-342. https://doi.org/10.1016/0301-0511(95)05165-1

Verwey, W. B., & Zaidel, D. M. (2000). Predicting drowsiness accidents from personal attributes, eye blinks and ongoing driving behaviour. *Personality and Individual*

Differences, *28*(1), 123-142. https://doi.org/https://doi.org/10.1016/S0191-8869(99)00089-6

vhp human performance. (2024). *Beoordeling werkdruk*. https://vhphp.sharepoint.com

von Gall, C., Muth, T., & Angerer, P. (2023). Sleep Duration on Workdays Is Correlated with Subjective Workload and Subjective Impact of High Workload on Sleep in Young Healthy Adults. *Brain Sciences*, *13*(5), 818. https://doi.org/10.3390/brainsci13050818

Warm, J. S., Dember, W. N., & Hancock, P. A. (1996). Vigilance and workload in automated systems. In *Automation and human performance* (pp. 183-200). CRC Press. https://www.researchgate.net/publication/254199951_Vigilance_and_workload_in_automated_systems

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors*, *50*(3), 433-441. https://doi.org/10.1518/001872008X312152

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. Routledge. https://doi.org/https://doi.org/10.4324/9781003177616

Wood, J. (2007). 16 - CCTV Ergonomics: Case Studies and Practical Guidance. In R. N. Pikaar, E. A. P. Koningsveld, & P. J. M. Settels (Eds.), *Meeting Diversity in Ergonomics* (pp. 271-287). Elsevier Science Ltd. https://doi.org/https://doi.org/10.1016/B978-008045373-6/50017-9

Yan, S., Tran, C. C., Chen, Y., Tan, K., & Habiyaremye, J. L. (2017). Effect of user interface layout on the operators' mental workload in emergency operating procedures in nuclear power plants. *Nuclear Engineering and Design*, *322*, 266-276. https://doi.org/https://doi.org/10.1016/j.nucengdes.2017.07.012

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of comparative neurology and psychology*, *18*(5), 459-482.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics*, *58*(1), 1-17. https://doi.org/10.1080/00140139.2014.956151

Young, M. S., & Stanton, N. A. (2004). Mental workload. In *Handbook of human factors and ergonomics methods* (pp. 416-426). CRC Press. https://ftp.idu.ac.id/wp-content/uploads/ebook/ip/BUKU%20ERGONOMI/BUKU%20INGGRIS/HumanFactors.pdf

Young, M. S., & Stanton, N. A. (2006). Mental workload: theory, measurement, and application. In W. Karwowski (Ed.), *International Encyclopedia of Ergonomics and Human Factors-3 Volume Set* (pp. 818-821). CRC Press. https://www.taylorfrancis.com/chapters/edit/10.1201/9780849375477-178/mental-workload-theory-measurement-application

Zanabazar, A., & Jigjiddorj, S. (2022). Relationships between mental workload, job burnout, and organizational commitment. *SHS Web Conf.*, *132*, 01003. https://doi.org/10.1051/shsconf/202213201003

Appendix A
Questionnaires

Figure A1

Questionnaire A1

NASA-TLX Dutch

Deelnemer- en Conditienummer Invoer

Page description:
Geef alstublieft uw deelnemer- en conditienummer op.

1. Conditienummer
*

2. Deelnemernummer *

Vragenlijst

In deze vragenlijst meten we jouw slaperigheid en mentale werkdruk. Je antwoorden zullen ons helpen begrijpen hoe deze factoren jouw recente ervaring hebben beïnvloed.

NASA-TLX: Instructies

Deze vragenlijst is ontwikkeld door NASA om de werkdruk te meten die je hebt ervaren tijdens het uitvoeren van de taak die je net hebt uitgevoerd. De vragenlijst gebruikt zes factoren om de werkdruk te beoordelen.

Deze zes factoren worden op de volgende pagina gedefinieerd. Lees ze zorgvuldig door om ervoor te zorgen dat je begrijpt wat elke factor betekent. Als je vragen hebt, stel deze dan aan de onderzoeker (Léon).

NASA-TLX: Definities

Mentale Belasting (Laag/Hoog)
Hoeveel mentale en perceptuele activiteit was vereist (bijv. denken, beslissen, rekenen, onthouden, kijken, zoeken, etc.)? Was de taak gemakkelijk of veeleisend, eenvoudig of complex, nauwkeurig of tolerant voor fouten?

Fysieke Belasting (Laag/Hoog)
Hoeveel fysieke activiteit was vereist (bijv. duwen, trekken, draaien, bedienen, activeren, etc.)? Was de taak gemakkelijk of veeleisend, traag of snel, ontspannen of inspannend, rustgevend of vermoeiend?

Tijdsdruk (Laag/Hoog)
Hoeveel tijdsdruk voelde je door het tempo waarin de taken of taakonderdelen plaatsvonden? Was het tempo traag en ontspannen of snel en hectisch?

Prestatie (Goed/Slecht)
Hoe succesvol denk je dat je was in het bereiken van de doelen van de taak die door de onderzoeker (of jezelf) waren gesteld? Hoe tevreden was je met je prestaties bij het behalen van deze doelen?

Inspanning (Laag/Hoog)
Hoeveel moeite moest je doen (mentaal en fysiek) om je prestatieniveau te behalen?

Frustratieniveau (Laag/Hoog)
Hoe onzeker, ontmoedigd, geïrriteerd, gestrest en geërgerd voelde je je versus hoe zeker, tevreden, ontspannen en voldaan voelde je je tijdens de taak?

Beoordelingsschalen

Page description:

Je krijgt nu een reeks beoordelingsschalen te zien.

Voor elk van de zes schalen evalueer je de taak die je zojuist hebt uitgevoerd door de plek op de schaal te selecteren die het beste overeenkomt met jouw ervaring. Elke schaal wordt beschreven aan de hand van de beschrijvingen op de twee uiteinden van de lijn.

Overweeg je antwoord zorgvuldig bij het onderscheiden van de verschillende taakomstandigheden en beoordeel elke schaal afzonderlijk.

Mentale Belasting

Page description:
Selecteer uw antwoord op de onderstaande schaal.

3. Hoe mentaal belastend was deze taak? *

Laag ——— Hoog

Fysieke Belasting

Page description:
Selecteer uw antwoord op de onderstaande schaal.

4. Hoeveel fysieke inspanning hebt u geleverd voor deze taak? *

Laag ——— Hoog

Tijdsdruk

Page description:
Selecteer uw antwoord op de onderstaande schaal.

5. Hoeveel tijdsdruk voelde u om deze taak te voltooien?
*

Laag | Hoog

Prestaties

Page description:
Selecteer uw antwoord op de onderstaande schaal.

6. Hoe succesvol denkt u dat u was in het bereiken van de taak?
*

Goed | Slecht

Inspanning

Page description:
Selecteer uw antwoord op de onderstaande schaal.

7. Hoeveel moeite moest u doen om uw prestatieniveau te bereiken?
*

Laag | Hoog

Frustratie

Page description:
Selecteer uw antwoord op de onderstaande schaal.

8. Hoe onzeker, ontmoedigd, geïrriteerd, gestrest en gefrustreerd voelde u zich tijdens deze taak? *

Laag Hoog

Karolinska Sleepiness Scale

Beoordeel uw huidige niveau van slaperigheid op de volgende schaal. Richt u alleen op hoe u zich op dit exacte moment voelt, niet op hoe u zich eerder voelde of hoe u denkt dat u zich later zult voelen.

9. Hoe alert of slaperig voelt u zich op dit moment? *

- Extreem alert
- Erg alert
- Alert
- Redelijk alert
- Niet alert maak oop niet slaperig
- Enkele tekenen van slaperigheid
- Slaperig, maar geen moeite om alert te blijven
- Slaperig, wat moeite om alert te blijven
- Erg slaperig, het kost veel moeite om wakker te blijven

Bekendheid met de incidenten

10. Heeft u eerder beelden van een van de incidenten die u zojuist heeft gezien? *

- Nee
- Ja

11. Indien ja, welke van de zojuist getoonde incidenten heeft u eerder gezien? Beschrijf het/deze kort in een paar woorden.

Dank u wel!

Page description:

Dank u wel voor het invullen van onze enquête. Uw antwoord is erg belangrijk voor ons.

Note. This questionnaire includes the Dutch translation of the NASA-TLX scale used to assess subjective mental workload. Additionally, a Dutch translation of the Karolinska Sleepiness Scale as well as a question investigating whether operators had seen this exact traffic footage previously during their work were integrated into the questionnaire.

Figure A2

Questionnaire A2

NASA-TLX Pairwise Comparisons

Deelnemersnummer

1. Wat is uw deelnemersnummer? *

Parenvergelijkingen

Page description:

U krijgt nu een reeks paren van beoordelingsschaalfactoren te zien; elk paar verschijnt op een apart scherm.

Kies bij elk paar de factor die belangrijker was voor uw beleving van de werkdruk in de taak die u zojuist heeft uitgevoerd.

Parenvergelijkingen

Page description:

2. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Inspanning
- Fysieke Belasting

Parenvergelijkingen

Page description:

3. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Tijdsdruk
- Mentale Belasting

Parenvergelijkingen

Page description:

4. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Frustratie
- Mentale Belasting

Parenvergelijkingen

Page description:

5. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Tijdsdruk
- Inspanning

Parenvergelijkingen

Page description:

6. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Frustratie
- Inspanning

Parenvergelijkingen

Page description:

7. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Mentale Belasting
- Inspanning

Parenvergelijkingen

Page description:

8. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Fysieke Belasting
- Frustratie

Parenvergelijkingen

Page description:

9. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Fysieke Belasting
- Tijdsdruk

Parenvergelijkingen

Page description:

10. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Fysieke Belasting
- Prestatie

Parenvergelijkingen

Page description:

11. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Prestatie
- Mentale Belasting

Parenvergelijkingen

Page description:

12. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Prestatie
- Tijdsdruk

Parenvergelijkingen

Page description:

13. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Inspanning
- Prestatie

Parenvergelijkingen

Page description:

14. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Prestatie
- Frustratie

Parenvergelijkingen

Page description:

15. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Mentale Belasting
- Fysieke Belasting

Parenvergelijkingen

Page description:

16. Kies hieronder de factor die de belangrijkste bijdrage leverde aan de werkdruk voor de specifieke taak die u zojuist heeft uitgevoerd. *

- Tijdsdruk
- Frustratie

Deelnemerinformatie

17. Wat is uw leeftijd? *

- Onder de 20
- 20-29
- 30-39
- 40-49
- 50-59
- 60 of ouder

18. Hoe lang werkt u al als operator?
Gelieve het in jaren en maanden te beantwoorden. *

19. Heeft u visuele afwijkingen (bijv. bijziendheid, verziendheid)? *

- Nee
- Ja

20. **Indien ja**, zijn deze afwijkingen gecorrigeerd (bijv. met een bril of contactlenzen) of niet?

- Gecorrigeerd
- Niet gecorrigeerd

Dank u wel!

Page description:

Dank u wel voor het invullen van onze enquête. Uw antwoord is erg belangrijk voor ons.

Note. This questionnaire includes the Dutch translation of the NASA-TLX pairwise comparison procedure needed to calculate weighted scores of each subscale. Additionally, small number of questions inquiring about demographic information are asked at the end of the questionnaire.

Appendix B

Information- and Consent Forms

Figure B1

Dutch Informed Consent From

Toestemmingsformulier voor het Experiment over Mentale Werkdruk in Verkeerscentrales

U ONTVANGT EEN KOPIE VAN DIT TOESTEMMINGSFORMULIER

Vink de juiste vakjes aan voor uw toestemming	**Ja**	**Nee**
Deelname aan het onderzoek		
Ik heb de informatie over het onderzoek, gedateerd op 24.10.2024, gelezen en begrepen, of deze is mij voorgelezen. Ik heb vragen kunnen stellen over het onderzoek en mijn vragen zijn naar tevredenheid beantwoord.	○	○
Ik geef vrijwillig toestemming om deel te nemen aan dit onderzoek en begrijp dat ik vragen mag weigeren te beantwoorden en op elk moment mag stoppen met het onderzoek, zonder opgaaf van reden.	○	○
Ik begrijp dat deelname aan het onderzoek het vastleggen van mijn demografische gegevens (door mij ingevuld) omvat, evenals informatie over mentale werkdruk in de vorm van schriftelijke vragenlijsten (door mij ingevuld), eye-trackinggegevens in de vorm van data verzameld met de Tobii Pro Glasses 2, en gedragsinformatie zoals reactietijden (vastgelegd door middel van een toetsaanslag).	○	○
Gebruik van de informatie in het onderzoek		
Ik begrijp dat de informatie die ik verstrek uitsluitend voor onderzoeksdoeleinden zal worden gebruikt, wat het schrijven van een masterthese inhoudt, en niet buiten het onderzoeksteam zal worden gedeeld. Geanonimiseerde gegevens die voortkomen uit de analyse van de data kunnen worden gedeeld met de partijen die betrokken zijn bij de gegevensverzameling, zoals vhp human performance, de gemeenten Rotterdam, den Haag en Amsterdam.	○	○
Ik begrijp dat informatie zoals mijn leeftijd of werkervaring geanonimiseerd zal worden en niet buiten het team dat betrokken is bij het afstudeerproject zal worden gedeeld. Na de voltooiing van het project, of uiterlijk binnen een jaar, zullen alle geanonimiseerde gegevens met betrekking tot leeftijd of werkervaring worden vernietigd.	○	○
Toekomstig gebruik en hergebruik van de informatie door anderen		
Ik geef toestemming voor de eye-trackinggegevens die ik verstrek om anoniem te worden gearchiveerd bij vhp human performance voor maximaal één jaar, en anoniem bij de betrouwbare opslagplaatsen van de Universiteit Twente voor minimaal 10 jaar. De gegevens zullen worden geanonimiseerd met behulp van een willekeurig gegenereerd deelnemernummer en er zal geen link zijn naar persoonlijke identificatie gegevens. De resultaten van het onderzoek zullen alleen in geanonimiseerde vorm worden gedeeld met de partijen die betrokken zijn bij de gegevensverzameling.	○	○
Ik geef toestemming voor de demografische gegevens en vragenlijstgegevens die ik verstrek om anoniem te worden gearchiveerd bij vhp human performance voor maximaal één jaar en anoniem bij de betrouwbare opslagplaatsen van de Universiteit Twente voor minimaal 10 jaar. De gegevens zullen worden geanonimiseerd met behulp van een willekeurig gegenereerd deelnemernummer en er zal geen link zijn naar persoonlijke identificatie gegevens. De resultaten van het onderzoek zullen alleen in geanonimiseerde vorm worden gedeeld met de partijen die betrokken zijn bij de gegevensverzameling.	○	○
Ik geef toestemming voor de gedragsgegevens die ik verstrek om anoniem te worden gearchiveerd bij vhp human performance voor maximaal één jaar en anoniem bij de betrouwbare opslagplaatsen van de Universiteit Twente voor minimaal 10 jaar. De gegevens zullen worden geanonimiseerd met behulp van een willekeurig gegenereerd deelnemernummer en er zal geen link zijn naar persoonlijke identificatie gegevens. De resultaten van het onderzoek zullen alleen in geanonimiseerde vorm worden gedeeld met de partijen die betrokken zijn bij de gegevensverzameling.	○	○

vhp HUMAN PERFORMANCE

UNIVERSITY OF TWENTE.

Handtekeningen

_____________________ _____________________ ________
Naam van deelnemer Handtekening Datum

Ik heb het informatieblad nauwkeurig voorgelezen aan de potentiële deelnemer en heb, naar beste vermogen, ervoor gezorgd dat de deelnemer begrijpt waar zij vrijwillig toestemming voor geven.

________________________ __________________ ________
Léon Schlimme Handtekening Datum

Contactgegevens
Contactgegevens van de onderzoeker voor verdere informatie:

Léon Schlimme
leonschlimme@vhp.nl

Contactgegevens van de begeleider van dit project bij vhp human performance:

Nienke Bierhuizen
nienkebierhuizen@vhp.nl

Contactgegevens voor vragen over uw rechten als onderzoeksdeelnemer
Als u vragen heeft over uw rechten als onderzoeksdeelnemer, of informatie wilt verkrijgen, vragen wilt stellen, of zorgen wilt bespreken over dit onderzoek met iemand anders dan de onderzoeker(s), neem dan contact op met de secretaris van de Ethiekcommissie/domein Humanities & Social Sciences van de Faculteit Gedrags-, Management- en Sociale Wetenschappen aan de Universiteit Twente via e-mail: **ethicscommittee-hss@utwente.nl**

Figure B2

Dutch Information Sheet

Informatie over het onderzoek

Inleiding
Verkeersleiders en/of tunneloperators werken vaak met veel camerabeelden ofwel CCTV-beelden. We weten dat de mens maar een beperkte hoeveelheid informatie of camerabeelden tegelijk kan bekijken en verwerken. Bijvoorbeeld: als je naar één camerabeeld kijkt is de kans veel groter dat je een afwijking meteen waarneemt dan wanneer je tien of misschien wel twintig camerabeelden tegelijk monitort. Dit komt omdat je je aandacht moet verdelen. Wanneer een operator te veel beelden tegelijk moet bekijken dan kan het zijn dat er informatie gemist wordt. Daarnaast kan deze grote hoeveelheid informatie zorgen voor cognitieve overbelasting, hoge ervaren werkdruk of stress. Er is niet eerder onderzoek gedaan naar hoeveel camerabeelden een tunneloperator tegelijk kan waarnemen, en wat het effect hiervan is op de prestatie en ervaren werkdruk van de operator.

Mijn naam is Léon Schlimme en ik studeer human factors & engineering psychology aan de Universiteit van Twente. Voor mijn afstudeeronderzoek onderzoek ik mentale werkbelasting in de context van controlekamers, met een specifieke focus op de verkeerscentrale. Verkeersleiders en/of tunneloperators werken vaak met veel camerabeelden ofwel CCTV-beelden. In mijn afstudeeronderzoek bij vhp human performance onderzoek ik hoe het aantal CCTV-beelden de prestatie en mentale werkdruk van operators beïnvloedt. Ik meet deze werkbelasting met behulp van een eye-trackingbril en enkele vragenlijsten, die in dit document verder worden toegelicht.

Doel van het onderzoek
Het doel van het onderzoek is om te kijken wat het effect van het aantal camerabeelden is op de prestatie en mentale werkdruk van tunneloperators. Daarnaast is het doel om te begrijpen waarom er wel of geen effect is. Bijvoorbeeld:

- Kan een tunneloperator beter een incident in een tunnel afhandelen met zestien of met vierentwintig camerabeelden? Waarom?
- Ervaart een tunneloperator meer of minder werkdruk of stress bij zestien of bij vierentwintig camerabeelden? Waarom?

De uitkomsten van dit onderzoek kunnen mogelijk gebruikt worden om in de toekomst werkplekken en verkeerscentrales te ontwerpen. Het doel is uiteindelijk om ervoor te zorgen dat de operators zo gezond en prettig mogelijk hun werk kunnen uitvoeren.

Risico's van deelname
Dit onderzoeksproject is goedgekeurd door de BMS Ethics Committee van de Universiteit Twente.

Er zijn geen andere risico's verbonden aan deelname aan dit onderzoek dan de risico's die verbonden zijn aan de dagelijkse verantwoordelijkheden van een operator in de verkeerscentrale.

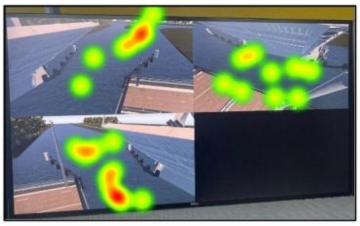
Procedure voor terugtrekking
U kunt op elk moment besluiten te stoppen met deelname aan het onderzoek, zonder dat u een verklaring hoeft te geven. Informeer de onderzoeker als u wilt stoppen.
Als een deelnemer zich terugtrekt, worden alle reeds verzamelde gegevens uitgesloten, indien gewenst. Er zullen geen verdere gegevens worden verzameld.

vhp HUMAN PERFORMANCE

UNIVERSITY OF TWENTE.

Persoonlijke informatie en gegevensgebruik
Deelname aan het onderzoek is volledig anoniem. Er worden wel demografische gegevens (zoals leeftijd en geslacht), eye-trackinginformatie, vragenlijsten over mentale werkdruk en reactietijden vastgelegd omdat dit relevant is voor het onderzoek. Deze gegevens zullen vertrouwelijk worden verwerkt en geanonimiseerd. Alle verzamelde gegevens zullen dus niet worden gekoppeld aan uw naam of andere identificerende gegevens. De gegevens zullen uitsluitend voor onderzoeksdoeleinden worden gebruikt, wat het schrijven van een masterscriptie inhoudt. Na de voltooiing van het afstudeerproject, dat naar verwachting vóór 01.02.2025 wordt afgerond, zullen alle gegevens over leeftijd en werkervaring worden verwijderd.
Geanonimiseerde onderzoeksgegevens kunnen worden gearchiveerd voor toekomstig onderzoek of gedeeld worden met de partijen die betrokken zijn bij de gegevensverzameling, zoals vhp human performance, de gemeenten Rotterdam, Den Haag en Amsterdam, maar kunnen niet worden herleid naar de deelnemers.
Deelnemers hebben het recht om toegang te krijgen tot hun gegevens, verzoeken om correcties van eventuele onnauwkeurigheden, of het verzoek om hun gegevens op elk moment te verwijderen. Als u besluit om u terug te trekken uit het onderzoek, kunt u ook verzoeken om de verwijdering van alle verzamelde gegevens van u.

Eye Trackingbril
Eye tracking is een methode die wordt gebruikt om te meten en te analyseren waar een persoon naar kijkt. De Tobii Pro Glasses 2 is een draagbare bril die gedetailleerde gegevens over oogbewegingen vastleggen, waardoor onderzoekers aandacht, visueel gedrag en cognitieve processen kunnen meten. De Tobii Pro Glasses 2 registreren verschillende soorten gegevens van deelnemers, waaronder kijkrichting, fixatiepunten (waar de persoon langere tijd naar kijkt), sacades (snelle oogbewegingen tussen fixatiepunten), pupilverwijding en de duur van elke blik. Deze informatie helpt te begrijpen hoe deelnemers interacteren met hun omgeving en kan worden gebruikt om hun focus, aandachtverdeling en mentale werkdruk te beoordelen tijdens het monitoren van camerabeelden. Er worden enkel eye-tracking opnames gemaakt van het bekijken van de camerabeelden.



Bewaartermijn van onderzoeksgegevens
Onderzoeksgegevens, die geanonimiseerde gegevens bevatten die voortkomen uit het onderzoek, zullen anoniem worden opgeslagen bij vhp human performance voor maximaal 1 jaar, en worden opgeslagen in betrouwbare opslagplaatsen voor onderzoeksgegevens van de Universiteit Twente voor een minimale duur van 10 jaar, in overeenstemming met het beleid voor Onderzoeksgegevensbeheer van de Universiteit Twente. De gegevens zullen alleen toegankelijk zijn voor geautoriseerde onderzoekers voor onderzoeksdoeleinden. De masterscriptie die voortkomt uit dit onderzoek zal worden opgeslagen in de repository voor studentenproeven van de Universiteit Twente en zal worden gedeeld met vhp human performance, de gemeenten Rotterdam, Den Haag, en Amsterdam.

vhp HUMAN PERFORMANCE

UNIVERSITY OF TWENTE.

Contactgegevens
Als u nog vragen heeft, kunt u contact met mij opnemen:
Léon Schlimme
leonschlimme@vhp.nl

Contactgegevens van de begeleider van dit project bij vhp human performance:
Nienke Bierhuizen
nienkebierhuizen@vhp.nl

Contactgegevens voor vragen over uw rechten als onderzoeksdeelnemer
Als u vragen heeft over uw rechten als onderzoeksdeelnemer, of informatie wilt verkrijgen, vragen wilt stellen, of zorgen wilt bespreken over dit onderzoek met iemand anders dan de onderzoeker(s), neem dan contact op met de secretaris van de Ethiekcommissie/domein Humanities & Social Sciences van de Faculteit Gedrags-, Management- en Sociale Wetenschappen aan de Universiteit Twente via e-mail: **ethicscommittee-hss@utwente.nl**

Appendix C
Task Instructions

Figure C1

Dutch Task Instructions

Instructies

Instructies voor de monitoringsopdracht

Uw taak voor dit onderzoek is om de camerabeelden te monitoren en **zo snel mogelijk** op de spatiebalk te drukken zodra u een incident opmerkt.

Hiervoor houdt u uw wijsvinger op de spatiebalk terwijl u de beelden monitort.

Incidenten kunnen het volgende omvatten:
• Voertuigen die stoppen in de tunnel of op de weg
• Ongelukken
• Ongeautoriseerde fietsers of voetgangers die een weg of tunnel betreden

Druk alstublieft slechts één keer per incident op de spatiebalk.

Let op: incidenten zijn niet altijd direct zichtbaar op de camera. Incidenten kunnen ook worden herkend aan signalen, zoals een plotselinge afname van het aantal voertuigen dat de tunnel binnenrijdt, of een opstopping van verkeer in de tunnel.

Gedurende het onderzoek zult u verschillende video's bekijken. Het is normaal dat de video scenario's elkaar afwisselen tijdens dezelfde ronde.

Tijdens het experiment zult u een eye-trackingbril dragen. Léon zal u hier meer over vertellen.

Zet uw telefoon op stille modus en vermijd afleidingen tijdens het onderzoek.

Het onderzoek bestaat uit drie delen van ongeveer 25 minuten. Na elk deel vult u een vragenlijst in.

Als u vragen heeft over het experiment, stel deze dan nu gerust.

Appendix D
Location-Specific Material

At the first location, a Panasonic FZ570 WUXGA projector with a resolution of 1920 x 1200 pixels and a refresh rate of 60 Hz, which was mounted to the ceiling above the participant, was used. The image was projected onto a white surface. At the second location, a Video Wall UDE-H Series screen by Samsung with a resolution of 1920 x 1080 pixels and a refresh rate of 60 Hz was used. At the third location, a Q-Line Display screen by Philips with a resolution of 3840 x 2160 pixels and a refresh rate of 60 Hz was used. In a fourth setup, a FHD standard signage screen by LG with a resolution of 1920 x 1080 pixels and a refresh rate of 60 Hz was used. More details are noted in Table D1.

Table D1

Location-specific Material

	A (cm)	B (cm)	C (in)	D (°)	E (°)	*n*
Setup 1	280	150	100*	43*	23*	3
Setup 2	130	118	55	50	27	6
Setup 3	150	122	65	51	28	6
Setup 4	145	124	55	46	24	1

Note. A describes the distance between operator and screen, B describes the distance between the floor and the bottom edge of the screen, and C describes the screen diagonal (Figure 3). The screen size and visual angles for Setup 1 marked with * were estimated using images of the setup, as the exact size was not measured. D represents the horizontal visual angle of the screen, and E represents the vertical visual angle of the screen from the participant's perspective. The number of participants per setup is represented by *n*.

Appendix E

Use of AI

During the preparation of this work the author used ChatGPT in order to brainstorm ideas for coding in R, to help debugging and to rephrase sentences that may have been confusing. After using this tool/service, the author reviewed and edited the content as needed and takes full responsibility for the content of the work.