I Spy with my AI: The Effects of AI-based Visual Cueing on Human Operators' Performance and Cognitive Load in CCTV Control Rooms

Jonas De Bruyne[a,b,*], Jamil Joundi[a,c], Jessica Morton[a,b], Aleksandra Zheleva[a,b], Niels Van Kets[d], Glenn Van Wallendael[d], Durk Talsma[e], Jelle Saldien[a,c], Lieven De Marez[a,b], Wouter Durnez[a,b] and Klaas Bombeke[a,b]

[a] imec-mict-UGent, Miriam Makebaplein 1, Ghent, 9000, Belgium
[b] Department of Communication Sciences, Ghent University, Sint-Pietersnieuwstraat 41, Ghent, 9000, Belgium
[c] Department of Industrial Systems Engineering and Product Design, Ghent University, Technologiepark 46, Zwijnaarde, 9052, Belgium
[d] imec-IDLab-UGent, Department of Electronics and Information Systems, Ghent University, Technologiepark 126, Zwijnaarde, 9052, Belgium
[e] Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, Ghent, 9000, Belgium

ARTICLE INFO

Keywords:
Virtual Reality
User Testing
Visual cueing
Cognitive Load
CCTV
Artificial Intelligence

ABSTRACT

The increased number of security cameras in modern cities has elevated the video-feed monitoring demands of closed-circuit television (CCTV) operators. As a result, new AI-driven support systems that leverage the power of computer vision algorithms have been deployed to facilitate the operators' work. However, to effectively design intuitive, AI-driven interfaces and validate their impact on the operators' performance, extensive user testing is required. To address this, we previously developed and tested a virtual reality (VR) control room that can be used to iteratively evaluate intelligent computer assistants and interfaces while operators are subjected to different cognitive load. In the present study, we use this VR environment and physiological markers (e.g., eye tracking measures) to investigate how AI-based visual cueing (i.e., pushing forward video streams on which detections are highlighted by rectangles drawn around targets) affects operator performance and cognitive load. Results suggest that support systems using such technology in a control room improve operators' performance and decrease their cognitive load, as reflected by changes in pupil dilation and subjective reports irrespective of induced cognitive load.

1. Introduction

Police services began exploring the idea of using video cameras as means of patrolling public spaces over fifty years ago. In the past few decades, closed-circuit television (CCTV) monitoring for security purposes has grown, resulting in an exponential increase of security cameras in public and private places (Hollis, 2019; Norris et al., 2004). Despite the development of modern computer vision technologies (see Sreenu and Durai, 2019), many control rooms still operate in a traditional fashion. Control room operators are taxed with monitoring a large mosaic comprised of different camera streams. In addition, it is not unusual to have more cameras than screens, whereby operators frequently switch between video streams. This monitoring style relies heavily on human operators that have cognitive limitations such as limited working memory capacity (Keval and Sasse, 2006) and the fact that sustained attention heavily strains one's cognitive resources (Warm et al., 2008). As a result, the increased surveillance demands provided by multiple cameras can promote poor detection performance (Stainer et al., 2017).

Given the aforementioned limitations of human operators, it appears inevitable that intelligent surveillance techniques such as computer vision algorithms will increasingly be employed to assist operators. One such intelligent system that aims to overcome the operator's cognitive limitations is one that automates aspects of the operator's task (Hodgetts et al., 2017). A major part of CCTV control room operators' tasks is proactive surveillance (Keval and Sasse, 2006), which involves scanning many video streams to visually detect anomalies. As described by Keval and Sasse (2006), operators usually don't use a specific strategy here. Rather, operators often scan through video feeds at random. Consequently, many events are at risk of being missed. On this point, assistive computer vision technology, which automatically highlights events that need inspection, would be beneficial. Additionally, the development of interfaces that display video feeds to facilitate the process of monitoring and switching between multiple feeds during a surveillance task can decrease operators' cognitive demands (e.g., Pelletier et al., 2015). Therefore, it is imperative to investigate the added value of such assistive technologies and –because such systems still necessitate a human element– their effects on operators. In addition to validating, such iterative tests will inform the development of future optimal user-centric systems and add to existing –yet non-exhaustive– control room design guidelines (e.g., Grozdanovic and Janackovic, 2018; Pikaar et al., 2015).

*Corresponding author
Jonas.DeBruyne@UGent.be (J. De Bruyne)
ORCID(s): 0000-0002-6077-6084 (J. De Bruyne); 0002-3437-1972 (J. Joundi); 0000-0002-2677-478X (J. Morton); 0000-0003-3478-4969 (A. Zheleva); 0000-0001-5495-2240 (N. Van Kets); 0000-0001-9530-3466 (G. Van Wallendael); 0002-3435-2317 (D. Talsma); 0003-2557-3764 (J. Saldien); 0001-7716-4079 (L. De Marez); 0001-8045-8801 (W. Durnez); 0003-2056-1246 (K. Bombeke)

1.1. Cognitive load

Although the concept of cognitive load is intuitively easy to understand, there is no consensus on how it should be defined, leaving us with an incoherent theoretical understanding (Van Acker et al., 2018; Young et al., 2015). One operational definition is provided by Young and Stanton (2001, pp. 507). They suggest that cognitive load reflects "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". In their definition, attentional resources are limited. Thus, when available attentional resources are fully allocated, a cognitive overload results which often hinders performance. Evidently, this outcome is undesirable in control room environments. Next to the mediating factor of inter-individual differences, this definition highlights the possibility of mediating cognitive load by manipulating task demands (e.g., complexity, temporal pressure) and external support. It therefore suggests that it is valuable to review the effects of both assistive technology and task demands on workers.

In a recent study by Van Acker et al. (2018), however, the authors formulated an implementable framework on cognitive load by disentangling the concept into its antecedents, defining attributes and consequences. The authors of this paper pointed out the misuse of the term *mediator* in the above-cited definition by Young and Stanton (2001). In fact, they claimed that *task demands* are more a predictor (i.e., an antecedent) than a mediator of mental load. On the other hand, *past experience* and *external support* operate as moderators according to this framework. As an example, assistive technology influences the effect of task demands on cognitive load.

The framework proposed by Van Acker et al. (2018) states clearly that cognitive work demands are predictors of cognitive load, and consequently, high cognitive load can have detrimental work-related consequences (e.g., lower and slower performance). Interestingly, external support is a factor that can moderate this influence, especially in the current context of operator support systems.

Indeed, external support by intelligent technology has been shown to positively affect control room operators' subjective load. For instance, Dadashi et al. (2013) examined the effects of automation accuracy and reliability on attention capacity and perceived cognitive load in a video monitoring task. Participants monitored a video stream from one camera while assisted by simulated automation. Their task was to detect a previously described actor (i.e., the target) when he entered a scene. The automated assistance system drew rectangles around (potential) targets in specific colors. The color of these rectangles resembled the system's confidence in the highlighted actor being the target. The reliability of the simulated automation was manipulated by changing the frequency of each color. As an example, in the unreliable confidence condition, half of the hits (i.e., the detected actor is the target for detection) were identified with the color indicating high system confidence, and the other half were presented as low system confidence. The results illustrated an advantageous effect of reliable automated assistance on load — when the system consistently identified hits and false positives with high and low confidence levels, respectively, participants reported reduced load. Participants also performed a secondary task while engaged in monitoring where they counted the number of people carrying backpacks in the video feed. Secondary task performance was an indirect indication of spare mental capacity which is inversely related to cognitive load. The findings suggested more spare mental capacity when participants were assisted by a reliable automated system.

Although this study provided valuable insights for future system development, it was limited by its lack of a non-assisted control condition. Furthermore, in the experiment participants only monitored a single video stream at a time — a far cry from real-life control room settings, characterized by dozens of concurrent video streams. The first aim of the present study is to bridge this gap by immersing participants in a virtual reality simulator of a fully equipped CCTV control room. Also, the cognitive load framework of Van Acker et al. (2018) indicates that the antecedents *task complexity* and *task demand* are likely to indirectly affect task performance, which can itself be moderated by external support. Therefore, the second aim of the present study is to investigate and demonstrate how support systems can be evaluated under different cognitive loads. Lastly, in the study of Dadashi et al. (2013), cognitive load was only measured subjectively. However, there are multiple possibilities to measure cognitive load objectively.

1.2. Measuring cognitive load

Cognitive load as a multi-dimensional construct (Young et al., 2015) has been assessed through a wide range of procedures. First, questionnaires such as the adapted versions of the NASA-TLX questionnaire (Hart and Staveland, 1988a) are regularly used to measure perceived cognitive load (e.g., Di Nocera et al., 2007; DiDomenico and Nussbaum, 2008; Grier, 2015). Second, researchers have studied a number of physiological correlates of cognitive load, from electrical brain activity (Antonenko et al., 2010) to electrodermal activity (Setz et al., 2009) and pupil dilation (see, van der Wel and van Steenbergen, 2018). Better still, Vanneste et al. (2020) demonstrate that cognitive load assessment accuracy increases with a multimodal approach that includes multiple measures.

Electrical brain activity is typically measured using electroencephalography (EEG), which is a non-invasive neuroimaging technique that measures electrical brain activity via electrodes placed on the scalp. The resulting EEG signal is composed of oscillations in multiple frequency bands (e.g., delta, theta, and alpha). Performing spectral power analysis on this continuously recorded data allows researchers to investigate the power of these oscillations at different frequency bands. For instance, oscillations in the alpha range (8-12 Hz) are highly pronounced when people are in a relaxed yet wakeful state, and while their eyes are closed. However, when people open their eyes (i.e., moving from a relaxed to an attentional state of alertness) alpha activity is suppressed (or desynchronized). Importantly, there is ample evidence that when cognitive load increases, there is a reduction in alpha power (most prominent in parietal areas) and an increase in theta activity (most prominent in frontal areas) (Antonenko et al., 2010; Brouwer et al., 2012; Klimesch, 1996; Sauseng et al., 2010). Moreover, it has been shown that these cognitive load markers can be measured in actual control room working conditions (Fallahi et al., 2016).

Furthermore, it has been illustrated that pupil dilation and blink rate (i.e., number of eye blinks per minute) correlate with cognitive load. More precisely, the human pupil has been shown to dilate and blink rate to decrease with increasing cognitive load (e.g., Gavas et al., 2017; Krejtz et al., 2018; Ledger, 2013; van der Wel and van Steenbergen, 2018; Zheng et al., 2012). However, the effects of cognitive load on blink rate require distinguishing between visual demands and mental activity. The effect of cognitive load on blink rate, when driven by increased visual demands (e.g., in a search task), can be explained by an automatic adaptation that inhibits blinking as this impedes visual information processing (Borghini et al., 2014; Wanyan et al., 2018). In contrast, cognitive tasks that require no visual processing induce a speed-up blink rate with increasing cognitive load (Magliacano et al., 2020; Recarte et al., 2008). Blink rate also increases during mental rehearsals (De Jong and Merckelbach, 1990), that is, silently repeating information that needs to be remembered. In this sense, conflicting effects are expected when both a visual and a mental task (e.g., an arithmetic task) are to be performed interchangeably. Another study illustrated that the blink rate is generally low when the visual load is high. During such states, cognitive load does not have an identifiable impact on blink rate (i.e., floor effect). Thus, blink rate seems only influenced by cognitive load when visual load is low (Chen and Epps, 2014). Because of these conflicting predictions on the relationship between cognitive load and blink rate, in this study, blink rate is not used as a marker of cognitive load. However, it is still measured to explore how blink rate changes in the manipulated conditions to understand the possibility and suitability of using blink rate as a cognitive load marker in future research regarding the current context.

1.3. Virtual reality as a research tool

To effectuate a multimodal approach to assess cognitive load during CCTV monitoring, immersive virtual reality (VR) appears a promising testing environment. In VR, people can watch and interact with an immersive virtual environment (VE) by means of a head-mounted display (HMD). Building VR simulators offers multiple advantages over classical approaches.

First, it is less time-consuming to build and evaluate operator support systems in VR than it is to construct fully operational support systems before they can be tested (see also, Oberhauser and Dreyer, 2017). Also, VEs allow implementating Wizard of Oz (Dahlbäck et al., 1993) prototyping approach very easily to test initial ideas without the need of developing automated systems. Therefore researchers can simulate automated systems by manually steering in-scene events so that the participant believes these events are occurring automatically.

Second, it is mostly impossible to occupy existing and operational control rooms for prototype testing and experimenting, because that would impede ongoing work for several hours or even multiple days. Given that most CCTV control rooms should be operational 24/7 to guarantee, for instance, general safety in cities, it would be rather inappropriate to shut down such control rooms for experimental tests.

Third, VR allows researchers to create an environment that is fully controlled. Researchers can alter the lighting conditions, background noise, the presence of colleagues, etc. This facilitates a rigorous and more ecologically valid investigation of the experimental manipulation effects (e.g., the addition of assistive technology in surveillance rooms) during numerous controlled circumstances. In sum, VR empowers researchers to find an ideal balance between ecological validity and the advantages of a controlled design.

Finally, since all interactions with the VE are recorded, researchers can benefit from a rich data set that describes the participants' behavior (e.g., performance on a task, physical interaction with another agent or object in VR etc.). Furthermore, state-of-the-art HMDs with built-in eye trackers continuously log eye-related indices. As an example, insights on participants' preferences (e.g., where do they look the most?) as well as indirect indicators of cognitive load (pupil dilation and blink rate) can be derived from this data. Additionally, a study by Tauscher et al. (2019) demonstrated that, with some minor modifications, it is possible to also combine EEG and VR. Moreover, researchers have already been able to discriminate between different levels of cognitive load using a classical n-back task in an interactive VR environment regardless of the increase in muscle tension and activity as a result of the interactive environment (Tremmel et al., 2019).

1.4. The present study

Given the ongoing change toward semi-automated CCTV control rooms and because such systems require a human element, it appears important that the use and effectiveness of (new) assistive technology is investigated. Evaluating the effectiveness of these systems –and having an adequate method to do so– is crucial not only for assessing their performance, but also for informing the development of future user-centric systems, expanding control room design guidelines and evaluating their effects on human operators.

This study investigates how two visual cueing techniques that can be used in automated camera selection systems influence operators' performance, cognitive load, and behavior in camera surveillance control rooms under different levels of work pressure. The support system highlights specific events to the operator in two ways. The first is to push the camera feed of interest to one of the operator's personal screens, which serves to watch one of the camera streams into more detail. The second approach is to have the system draw rectangles around the person of interest in the camera stream that is pushed forward (Figure 1).

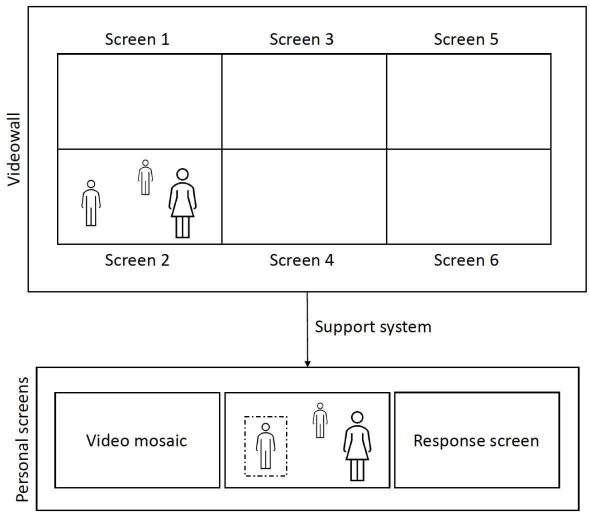

Figure 1: Flow of support system. The video of interest is pushed to one of the operators' personal screens by the support system and there a rectangle is drawn around the target.

In a previous study, a virtual CCTV control room was developed, where an operator's job was simulated and work pressure was manipulated (De Bruyne et al., 2021). Participants wore an HMD to interact with the virtual control room. They were asked to perform a simplified monitoring task (primary task) and from time to time were interrupted by auditory requests that required a response (secondary task). The secondary task either consisted of low demanding task rules and long response-stimulus intervals (RSIs) or high demanding task rules and short RSIs. The results suggested that manipulating these two features (task difficulty and RSIs) altered cognitive load as measured by a subjective measure (i.e., NASA-TLX questionnaire) and physiological markers (e.g., eye-tracking). Therefore, the secondary task manipulation of (De Bruyne et al., 2021) offers the opportunity to investigate the effect of new support systems while operators are under different cognitive loads.

The present study uses the same virtual control room and dual-task manipulation. The primary task in the present study, however, consists of an actual video monitoring task. While performing this task, participants are either assisted by the support system outlined above, or are not assisted at all. Given the general purpose of such support systems (i.e., increasing performance), the researchers investigate not only the influence of AI assistance on cognitive load but also on primary task performance. Furthermore, the interplay between the work pressure manipulation (i.e., current cognitive load levels) and the addition of the assisting technology is explored. In other words, higher task performance and lower experienced cognitive load are expected when participants are assisted by the visual cueing system. These outcomes are expected to be visible in the pupil size data (larger pupil size when not assisted), the EEG data (i.e., less alpha power when not assisted), and the subjective data (i.e., higher reported cognitive load when not assisted). Finally, we investigate whether an interaction exists between cognitive load and the assistive technology used (e.g., stronger effects of support system in high load conditions).

In sum, the present study aims to investigate the potential influence of an operator support system on the relationship between task performance, behavior, and cognitive load.

2. Method

2.1. Participants

31 participants (11 male, $M_{age} = 24.56$, $SD_{age} = 2.93$) took part in the experiment. Inclusion criteria were the type of hair (e.g., no dreadlocks as it interferes with EEG recordings) and history of simulator sickness (i.e., applicants who often suffer simulator sickness were not invited to the lab). Each participant signed informed consent and received 20 euros for their participation. The protocols of this study were approved by the Ethical Committee of the Faculty of Political and Social Sciences of Ghent University.

2.2. Materials and equipment

The VR setup consisted of a computer running SteamVR (v.1.14.16) and an HTC VIVE Pro Eye. The HMD and the controllers were tracked by two Vive SteamVR Base Stations 2.0. A wireless connection between the computer and the HMD using a Vive Wireless module (www.vive.com) was used to reduce the number of cables because both the HMD and an EEG cap were placed on the participant's head. The HMD's built-in eye-tracker and the Vive Eye-tracking Software Development Kit (SDK) SRanipal were used to obtain eye-tracking measures (incl. pupil sizes). This built-in eye-tracker had a sampling rate of 120 Hz. However, in this experiment, the eye-tracking data was recorded at frame rate (i.e., 50-60 Hz).

The experiment was built in Unity (version 2019.4.3f1) using the VRTK framework (vrtoolkit.readme.io/) for in-scene interactions. The VR environment was a pre-existing police control room asset that was modified according to the experiment's needs. The eventual virtual control room was equipped with a video wall consisting of 8 large screens and two desks with 3 monitors each (figure 2). One of the operator's (i.e., the participant) personal workspace monitors was used as a response screen with buttons that could be pressed using a pointer and the trigger button of the controller (which participants held in their hand of preference). Also, a walkie-talkie radio was present in the scene which was the audio source for the presentation of the auditory stimuli.

During experimental blocks, the six leftmost screens of the video wall each rendered one of the six camera angles while the two rightmost screens presented irrelevant news shows. One of the three personal screens of the operator showed a mosaic of all the six camera angles (see figure 2). The screen that was positioned in the middle of the operator's desk rendered one of the six videos in an enlarged format in order for the participant to see one of the videos on a more detailed level. When participants were not supported by the system, they could choose which of the six camera angles they wanted to see on that screen by clicking on one of the videos in the mosaic on the left screen. In contrast, this selection was done by the algorithm during AI-assisted conditions. In other words, during those conditions, the video selection montage was presented on the middle screen. Throughout the whole experiment, the artificially simulated luminance of the virtual control room was kept constant.

Figure 2: Overview of the virtual control room with displayed videos.

The video footage consisted of simulated camera surveillance videos that were shot in a bicycle storage room from six different angles. All angles were filmed using Go-Pro cameras (2x Go-Pro hero3, 2x Go-Pro hero7, and 2x Go-Pro hero5 session). Altogether, these cameras covered the entire bicycle storage. Four video sets were made of 6.5 minutes each. In the scene, actors simulated criminal events such as bicycle thefts and robberies. These events were spread randomly over the duration of each video. Next to criminal events, normal events such as people picking up their own bicycle, parking a bicycle or just passing by occurred frequently.

The video selection montage that was shown in the AI condition was created by Robovision (Belgium, robovision.ai) using an algorithm that detects people in a scene (based on RetinaNet; Lin et al., 2017). At any moment, the video in which most detections (i.e., people) were flagged by the algorithm was selected. A background subtraction approach, inspired by MOG2, was used to reduce false positives (Zivkovic, 2004). As a result, the algorithm ignored camera angles where there was a detection, yet no movement. Additionally, a temporal low pass filter was applied to the stream selection. This avoided very fast switches between different streams, and thus gave the operator time to interpret the content of the selected stream. Detections in the resulting video sequence – a montage that continuously switches between the six angles – were indicated by a green rectangle that was drawn around the detected person (figure 3). The video montage that represented the AI assistance was made so target events could not be missed if only the automatically selected stream was looked at. Importantly, the algorithm used in this study does not represent a commercially viable product. It was chosen to fit this experiment's goal, i.e., to investigate the influence of a system pushing forward one of many video streams on the operator's performance and cognitive load.

Figure 3: Example of video stream pushed by the algorithm with a highlighted detection (rectangle).

The EEG data were recorded using eegoTM mylab software (version 1.7.1; ANT Neuro, Netherlands, www.ant-neuro.com). WaveguardTM actively shielded caps (ANT Neuro, Netherlands, www.ant-neuro.com) with 64 Ag/AgCl electrodes placed according to the 10-20 system were used. These caps prevent 50/60 Hz environmental noise and artifacts arising from the movement of the cables in the recorded data. Recordings were referenced online to the CPz site electrode and the ground electrode was placed on AFz. The sampling frequency during recording was 512 Hz and impedance was kept below 25kΩ.

The perceived load was assessed at the end of every block using an adapted version of the NASA-TLX (Hart and Staveland, 1988b). The NASA-TLX is a well-known assessment instrument that indicates perceived load on six domains of task requirements (e.g., mental demand, physical demand, etc.).

2.3. The dual-task paradigm

2.3.1. The primary task

The primary task was actual video monitoring. Participants monitored the surveillance camera footage and were asked to report criminal events like bicycle thefts or robberies. They were instructed to press the 'detected' button on the response screen to flag an incident. When the participant pressed the 'detected' button, a timestamp was logged. These timestamps were used to assess performance on the primary task.

2.3.2. The secondary task

The currently used secondary task followed the outlines of the one illustrated in De Bruyne et al. (2021). Thus, it consisted of an auditory presentation of a sequence of randomly selected single-digit numbers ranging from 1 to 6. The length of each sequence varied from 2 to 6 digits at the trial level. In the low demand condition participants responded by clicking the last digit of the heard sequence on the response screen and then pressing the 'send' button. In contrast, the high demand condition imposed task rules dependent on the number of digits in the sequence. When the number of digits in the sequence was odd, participants clicked the last two heard digits and then press the 'send' button. When the number of digits in the sequence was even, participants clicked the first two digits they heard followed by the 'send' button. Moreover, the secondary task differed in RSI over both conditions. In the low demand condition, the RSI varied ad random between 25 and 30 seconds whereas in the high demand condition the RSI varied between 2 and 7 seconds. Reaction times as well as accuracy were measured.

2.4. Procedure

Upon arrival, participants signed informed consent and were briefed on how the EEG cap would be mounted and installed. After mounting the cap and adding conductive gel for the electrodes to have ideal impedance, the controller buttons and their usage were explained followed by a general introduction to the experiment. Participants wore the HMD on top of the EEG cap and held one controller in their hand of preference. Participants were clearly instructed that the primary task should be prioritized. Further detailed instructions on the tasks were presented onscreen in VR. Before the start of the experiment, baseline EEG data was recorded for 1 minute. During this minute, participants were asked to relax as much as possible with their eyes open. Subsequently, an eye-tracker calibration procedure was performed. After mounting the HMD, participants were allowed to get used to being in VR and familiarise themselves with the controller. Each participant performed every condition of the 2x2 design. Each of these conditions corresponded to one block. Participants were divided into four groups based on their subject number. Each group had a different combination of the order of secondary task demands conditions and the assistance condition (see, Figure 4). The condition randomization was generated following a Latin square design. The secondary task was practiced before the start of the experiment and before the start of the third block because, after the second block, the task rules of the secondary task changed. In other words, a participant either first performed two block with the task rules for the low load condition and subsequently two blocks with the task rules of the high load condition, or vice versa. After each block, participants answered the questions of the adapted version of the NASA-TLX questionnaire orally while they were still wearing the HMD.

	Block 1	Block 2	Block 3	Block 4
Group 1	manual - low	AI - low	manual - high	AI - high
Group 2	AI - low	manual - low	AI - high	manual - high
Group 3	manual - high	AI - high	manual - low	AI- low
Group 4	AI - high	manual - high	AI - low	manual - low

Figure 4: Overview of randomization of condition order.

2.5. Data analysis

The data was analyzed using two within-subject factors (i.e., *load* and *assistance* - high or low and manual or AI, respectively). However, as participants were only able to click on the mosaic screen in the manual condition, the dependent measure *number of clicks* (i.e., interactions with the interface) was analyzed using only *load* as a factor. All data pre-processing was performed in Python 3 and linear mixed-effects models (LMMs) were constructed in R using the lme4 package (Bates et al., 2014) specifying a random intercept for each participant. Degrees of freedom for the LMMs were corrected using Kenward-Rogers correction (Kenward and Roger, 1997). Before each analysis, outliers were removed from the data set. Outliers are defined as datapoints lower than -1.5 times the interquartile range (IQR) or higher than 1.5xIQR.

2.5.1. *Performance on the primary and secondary task*

Performance on the primary task was scored manually. Each 'detected' button press was categorized as either a *true positive* (TP) or a *false positive* (FP). Next, the criminal events that participants did not detect were counted and labeled as *false negative* (FN). The remaining non-criminal events that occurred in a block were counted as well and were labeled *true negative* (TN). Using the absolute counts of all TPs, TNs, FPs, and FNs during each block, two different measures were calculated for accuracy on the primary task for each condition: sensitivity (true positive rate; TPR) and specificity (true negative rate; TNR). Sensitivity (TPR) represents the probability with which a criminal event is detected as such. In contrast, specificity (TNR) represents the probability that a truly non-criminal event is marked as such. In other words, the proportion of non-criminal events that the operator correctly identified as non-criminal. In analytical terms, TPR is given by the number of TPs divided by the total number of *positive* events. Likewise, TNR is calculated by dividing the number of TNs by the total number of *negative* events.

To illustrate the concepts of sensitivity and specificity, imagine two operators, operator A and operator B. Operator A was able to detect all criminal events during a monitoring task. When he was in doubt, however, he flagged the doubtful event as criminal activity. By doing so, he decreased the probability of missing events (false negatives) but increased the probability of wrongfully labeling non-criminal events as criminal (false positives). Operator B was more conservative during the task. He marked events as criminal only when he was 100% sure. As a result, he often failed to detect criminal events (false negatives), but the probability of wrongfully judging an event as criminal (false positives) decreased. When comparing the operators, operator A would score better on sensitivity while operator B would score better on specificity.

Other than the measures described below, primary task performance was analyzed using a non-parametric test for repeated measures because the normality assumption did not hold given the nature of the data (i.e., skewed due to a ceiling effect). These models were built using the package nparLD (Noguchi et al., 2012) in R.

Performance on the secondary task was scored automatically during the experiment. If the response provided by the participant matched the correct response, the trial was scored 1. If the participant's response did not match the correct response, the trial was scored 0. The score on each trial was summed and divided by the total amount of trials within a block. Finally, this summed score was converted to a percentage. Reaction times on the secondary task were not analyzed as the different task rules across conditions required different responses.

2.5.2. Perceived cognitive load.

Responses on the adapted version of the NASA-TLX were scored on a scale ranging from 0 to 100. As has often been done in previous research, the mean of the responses across items was calculated resulting in one score of subjective cognitive load per participant.

2.5.3. Eye-tracking

Due to technical issues, eye-openness and pupil size were only recorded for 18 participants. However, the blink rate could also be calculated based on the EOG data (included in the EEG recording) instead of using the eye-openness data captured by the HMD. As a result, the mean blink rate (blinks/minute) was calculated per block for 30 participants. This calculation was performed using the neurokit2 package (Makowski et al., 2021).

Pupil size was logged at frame rate (50-60 Hz). This data was pre-processed by interpolating outliers (i.e., values that were smaller or larger than – 3 SD or larger than + 3 SD) and missing values due to blinks linearly on the subject level. Next, the mean pupil size for each participant for each condition was calculated.

As for eye gaze, three participants were excluded because their data was not captured continuously, resulting in critical data loss. The dependent variable that was calculated for eye gaze was time spent looking at the monitor rendering one enlarged video (i.e., the video pushed by the algorithm in the AI assistance condition) divided by the sum of the time spent looking at that monitor, and the time spent looking at the mosaic or the videowall (excluding the distractor video feeds on the video wall).

2.5.4. EEG data

The recorded EEG data were re-referenced to the average across all electrodes and filtered using a band-pass filter of 1-50 Hz. Ocular artifacts (blinks and eye movements) were isolated and removed from the continuous EEG data using independent component analysis (ICA) and by comparing these to the EOG data and, to confirm, after visual inspection of the components for each participant. Bad electrodes (drifting or flat-lining electrodes) were interpolated. Following, the data were resampled to 100 Hz and segmented per block before spectral power analyses were performed. Power in the alpha band and theta band was averaged across central parietal electrodes (CP1, CPz, CP2, P1, Pz, P2) and frontal electrodes (F1, Fz, F2, AF3, AFz, AF4) respectively and was calculated using Welch's method for spectral density estimation with a moving window of 2 seconds and 50% overlap. Also, a cognitive load index was calculated based on the theta Fz/alpha Pz ratio (Holm et al., 2009). This measure is another method used to assess cognitive load through EEG. The resulting dataset included alpha and theta power for the above-mentioned regions of interest and a cognitive load index. The dataset was analyzed using LMMs with a random intercept for each participant. For the analysis of the EEG data, one participant was excluded due to recording failure.

3. Results

3.1. Performance on the primary task.

For TPR, there was neither a main effect of assistance, $F(1, \infty) < 0.01$, $p = 0.986$, or load, $F(1, \infty) = 1.16$, $p = 0.282$, nor was there an interaction effect between load and assistance, $F(1, \infty) < 0.01$, $p = 0.978$. For TNR, there was a main effect of assistance, $F(1, \infty) = 3.95$, $p = 0.046$. This effect shows that TNR increased when participants were supported by the AI assistant (see, figure 5). Additionally, TNR was not influenced by load, $F(1, \infty) = 2.78$, $p = 0.096$, and there was no interaction effect between assistance and load, $F(1, \infty) = 0.94$, $p = 0.331$. The main effect of assistance on TNR illustrates that participants made fewer false positive errors when they were assisted by AI.

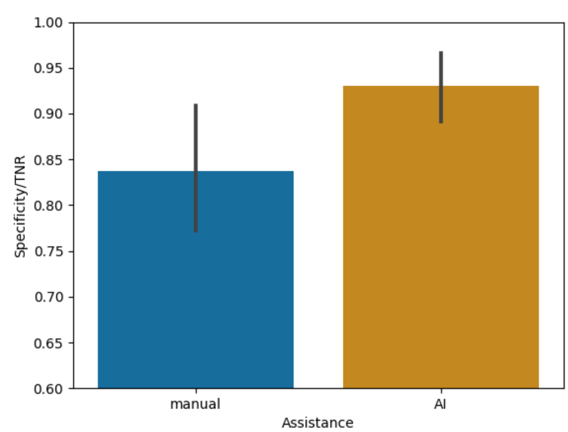

Figure 5: Average score on specificity by assistance condition. Error bars represent 95 CI.

3.2. Performance on the secondary task.

A main effect of load was found for accuracy on the secondary task, $F(1, 90) = 179.55$, $p < 0.001$, $\eta_p^2 = 0.67$, 95% CI [0.56, 0.74]. Finding such a convincing large effect size again serves as a positive manipulation check as participants performed better in the low load condition (M = 94.20%, SD = 6.80%) relative to the high load condition (M = 67.20%, SD = 17.10%). No main effect for assistance, $F(1, 90) = 0.433$, $p = 0.449$, $\eta_p^2 < 0.01$, 95%

CI [0, 0.08], and no interaction between load and assistance was found, $F(1, 90) = 0.43$, $p = 0.515$, $\eta_p^2 < 0.01$, 95% CI [0, 0.07].

3.3. Interactions with the navigation system in manual condition.

Load condition significantly affected the amount of clicks participants performed on the mosaic screen during the manual condition, $F(1, 27) = 22.09$, $p < 0.001$, $\eta_p^2 = 0.45$, 95% CI [0.17, 0.64]. Specifically, participants clicked less on the mosaic screen to look at one of the video feeds in more detail during the high load condition (M = 54.93, SD = 32.42) compared to the low load condition (M = 80.93, SD = 46.42). This large effect suggests that there were less mental resources available to actively scan through the video streams when cognitive load induced by the secondary task was high.

3.4. Eye gaze.

A main effect of assistance was found for the proportional time spent looking at the enlarged video screen, $F(1, 77) = 4.41$, $p = 0.039$, $\eta_p^2 = 0.05$, 95% CI [0, 0.18] (figure 6). This small to moderate effect shows that participants spent more time looking at the enlarged video feed when the AI assistant pushed one of the video feeds to the monitor in the middle of their desk (M = 67.85%, SD = 22.29%) relative to when they had to manually select video feeds to watch them in detail on that same monitor (M = 61.97%, SD = 19.57%). Inversely, this also means that participants visually explored the other video feeds presented on the mosaic screen and the video wall to a lesser extent when they were supported by the algorithm. For this measure, there was no main effect of load, $F(1, 77) < 0.03$, $p = 0.873$, $\eta_p^2 < 0.01$, 95% CI [0, 0.04], and no interaction effect between load and assistance, $F(1, 77) = 0.435$, $p = 0.512$, $\eta_p^2 < 0.01$, 95% CI [0, 0.08].

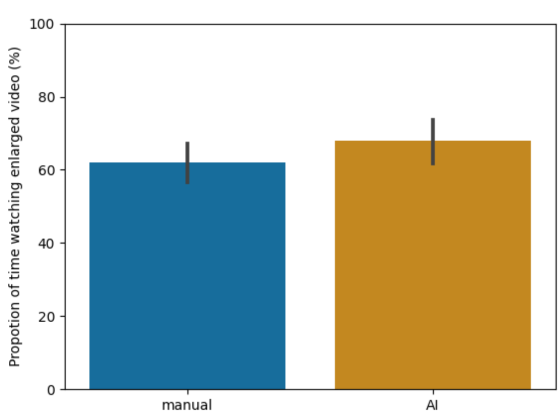

Figure 6: Average score on the proportion of time spent looking at the enlarged video by assistance condition. Error bars represent 95 CI.

3.5. Cognitive load.

Cognitive load was assessed using one subjective and three physiological measures. A main effect of load was found in the analysis of responses on the adapted version of the NASA-TLX questionnaire, $F(1, 84) = 108.79$, $p < 0.001$, $\eta_p^2 = 0.56$, 95% CI [0.43, 0.67]. This large effect indicates that the load manipulation affected perceived cognitive load. As such, the perceived cognitive load was higher in the high load condition compared to the low load condition. Next, a (moderate) main effect for assistance was observed, $F(1, 84) = 8.34$, $p = 0.005$, $\eta_p^2 = 0.09$, 95% CI [0.01, 0.22]. Participants' cognitive load increased with increasing load and was lower when they were supported by the AI system (figure 7). No significant interaction between load and assistance was found, $F(1, 84) = 0.74$, $p = 0.392$, $\eta_p^2 < 0.01$, 95% CI [0, 0.09].

The same main effects were found for pupil size. Specifically, participants' mean pupil size was higher when load was high (M = 3.51, SD = 0.35) compared to when load was low (M =3.45, SD = 0.35), $F(1, 48) = 7.08$, $p = 0.011$, $\eta_p^2 = 0.13$, 95% CI [0.01, 0.31]. Also, pupil sizes were larger when participants were not supported by AI (M = 3.52, SD = 0.36) compared to when they were (M = 3.45, SD = 0.35), $F(1, 48) = 8.31$, $p = 0.006$, $\eta_p^2 = 0.15$, 95% CI [0.01, 0.33]. Similar to what was found in the subjective measure of cognitive load, there was no interaction between load and assistance for pupil size, $F(1, 48) = 1.05$, $p = 0.310$, $\eta_p^2 = 0.02$, 95% CI [0, 0.16]. Next, blink rate also showed a main effect of load, $F(1, 51) = 5.21$, $p = 0.027$, $\eta_p^2 = 0.09$, 95% CI [0, 0.26]. Blink rate was higher in the high load condition (M = 12.86, SD = 9.03) relative to the low load condition (M = 10.76, SD = 7.65). In contrast to the two previous measures, however, no main effect of assistance on blink rate was found, $F(1, 51) = 3.41$, $p = 0.071$, $\eta_p^2 = 0.06$, 95% CI [0, 0.22]. Additionally, in parallel with perceived cognitive load and pupil size, no interaction between load and assistance was found for blink rate, $F(1, 51) = 0.11$, $p = 0.934$, $\eta_p^2 = 0.06$, 95% CI [0, 0.04].

As for the EEG data, alpha power in central parietal regions, theta power in frontal regions and the cognitive load index were analyzed. There was no main effect of load, $F(1, 87) = 3.74$, $p = 0.057$), $\eta_p^2 = 0.04$, 95% CI [0, 0.15], and no main effect of assistance on alpha power, $F(1, 87) = 3.36$, $p = 0.070$, , $\eta_p^2 = 0.04$, 95% CI [0, 0.14]. Also, no interaction between load and assistance was found, $F(1, 87) = 0.18$, $p = 0.674$, $\eta_p^2 < 0.01$, 95% CI [0, 0.06]. An additional analysis investigating the alpha power in the lower alpha range (8 Hz – 10 Hz) also showed no significant effects ($p > 0.05$). Theta power in frontal electrodes did not show a main effect of load, $F(1, 81) = 3.66$, $p = 0.059$, $\eta_p^2 = 0.04$, 95% CI [0, 0.16], nor a main effect of assistance, $F(1, 81) = 0.31$, $p = 0.581$, $\eta_p^2 < 0.01$, 95% CI [0, 0.08], nor an interaction between load and assistance, $F(1, 81) = 0.19$, $p = 0.667$, $\eta_p^2 < 0.01$, 95% CI [0, 0.06]. Similar to the findings on parietal alpha power and frontal theta power, no main effects of load, $F(1, 81) = 2.87$, $p = 0.094$, $\eta_p^2 = 0.03$, 95%

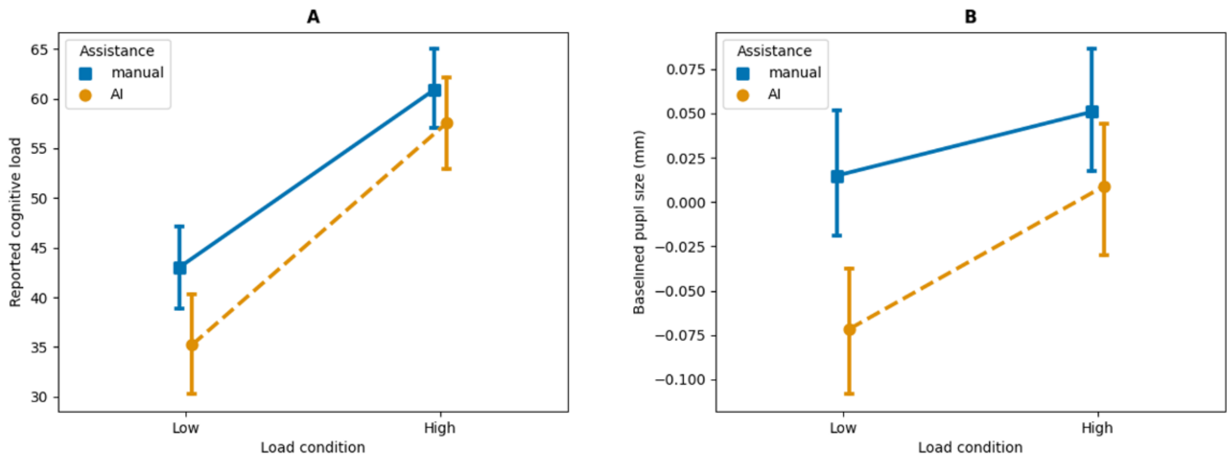

Figure 7: Mean scores on NASA-TLX questionnaire (A) and pupil size (B) across participants. For plotting purposes, baselined pupil size values are shown here. In the LMM, raw measures were used and individual variability was captured by a random intercept. Error bars indicate 95% confidence intervals.

CI [0, 0.14], and assistance, $F(1, 81) = 0.47$, $p = 0.497$, $\eta_p^2 < 0.01$, 95% CI [0, 0.08], were found for the cognitive load index. Also, the interaction between load and assistance was not significant, $F(1, 81) = 0.27$, $p = 0.605$, $\eta_p^2 < 0.01$, 95% CI [0, 0.07].

4. Discussion

In the current study, the researchers investigated the influence of two visual cueing techniques, employed in parallel, on operators' performance, behavior, and cognitive load in CCTV surveillance rooms. Because of the varying circumstances regarding cognitive load that are encountered by operators in control rooms, high and low load situations were simulated to gain insight into possible interactions between the operators' cognitive state and the (dis)advantageous effects of AI-based support systems. The virtual environment of a previous study (De Bruyne et al., 2021) was used, in which cognitive load was manipulated using a secondary task manipulation. The same manipulation was implemented in the current experiment to investigate the effects of the AI-based support system on performance, behavior, and cognitive load during different working circumstances (i.e., the operator experiences high or low cognitive load).

The effects of the current AI-based support system are promising. Results showed advantageous effects on operators' performance when helped by the intelligent computer assistant, as indicated by TNR. In-depth, this increase in performance seemed largely due to a decrease in false positive detections. This means that operators that are in fact binary classifying events as being criminal or non-criminal are less prone to judging an event that is truly non-criminal as criminal. This is advantageous because this means that resources will not be spent on false positive alarms. Think of a patrol unit that would be sent out to inspect the event in the field. However, one of the purposes behind offering assistive technology in CCTV control rooms is to increase sensitivity (i.e., TPR). In other words, to decrease the probability that target events are missed and thus increase the probability that all target events are detected. This effect was not found in the current study using the described support system. It remains possible, however, that such an effect would have been found if there would have been more than six different video streams in the experiment, because in real life, operators have to monitor an enormous amount of video streams and usually they randomly scan through the video streams (Keval and Sasse, 2006).

Next to an increase in performance, the rather large main effect of assistance on both subjective reports and pupil size strongly suggested a decrease in overall cognitive load when participants were assisted by the support system. This means that operators can benefit from assistive technology as more mental resources become available when they are assisted by the support system. Strikingly, no interaction effect between induced cognitive load and the presence of assistive technology was found. The findings on cognitive load as reflected by pupil size and the subjective reports, however, were inconsistent with the results of the other cognitive load markers. As an example, no main effect of assistance was found for performance on the secondary task. When the load elicited by the primary task decreases, one would expect an increase in performance on the secondary task (due to spare mental capacity). This would mean that this behavioral indicator of cognitive load was not influenced by the assistance manipulation. Explicitly requesting participants to prioritize the primary task, however, might have biased this finding.

Also for blink rate, even though the underlying drivers of the effect are still unknown, an effect that suggested a decrease in cognitive load when AI support was provided was not found. If increased blink rate could be attributed to increased mental activity and thus cognitive load, a decrease in blink rate when assisted by a support system would be expected. However, if the increased blink rate reflected the use of mental rehearsal (De Jong and Merckelbach, 1990) as a strategy to complete the secondary task, an effect of assistance would not be expected. Therefore, the absence of a main effect of assistance on blink rate does not yield concerns, except for the use of blink rate as a potential marker for cognitive load in the current experimental context. A solution for future research would be to include a secondary task for which it is impossible or not necessary to use mental rehearsal as a strategy to increase performance.

As for the neural markers of cognitive load in the current experiment, the measures reported might have been insensitive to changes in cognitive load given the highly demanding dual-task paradigm, including a primary task that required constant vigilance and active search. It has been demonstrated in the past that neural markers of cognitive load can reach a plateau after which cognitive load changes are no longer visible (Puma et al., 2018). As a consequence, and from what we learned from this study, it might be undesirable to include EEG measurements as a marker for cognitive load in the current and in similar contexts.

In sum, although there are inconsistencies in the findings regarding cognitive load as measured by different approaches, these inconsistencies might be the result of experiment-specific features and limitations which served the purpose of making this experiment as ecologically valid as possible.

Notably, this study did not control for every possible moderator from the implementable framework formulated by Van Acker et al. (2018) that might interplay with the effects of support systems on performance and cognitive load. One of these moderators, for instance, is experience. Future research should, therefore, consider including expert control room operators in the sample, as experts may interact differently with support systems or they may have developed different surveillance strategies over time. Other possible moderators such as job autonomy and visuospatial intelligence could also be considered in future experimental design reviews.

Using VR in this study was relatively new as compared to previous work in the field. There are, however, some concerns regarding generalisability to real-life control rooms when using a simulator as potential confounds – other than the simulator not being an exact copy of the real-world setting – that haven't been thought of for this study might have had an impact on the operators' behavior. Therefore, the results of this study have to be replicated, if practically possible, in existing surveillance rooms with professional control room operators, should strong claims about the effectiveness of the presented or similar operator support systems be made.

Because this study employs a novel methodology, we highlight some key takeaways regarding the use of VR in the current context both regarding the inclusion of the manipulation of cognitive load and the testing of new support systems in practice. First, it is important for the participants to familiarise themselves with the virtual environment and the controller(s). As VR remains novel for many people, providing some time for the participant to look around and try out in-scene interactions at the beginning of the experiment is highly recommended to avoid potential confounds induced by the novelty aspect. In this sense, it is especially desirable to include practice blocks in VR. Secondly, when simulating a surveillance task in VR, typical ways to control computer systems will have to be translated to a VR setting. Consequently, these translations have to be carefully considered to minimize the gap between operating the task in the simulator and in real life. Thirdly, most of the currently available HMDs have a limited field of view that is considerably different from what we are used to in real life. In a VR control room, this means that, for instance, fewer displays are positioned in an operator's peripheral field of view. This might result in less attention being paid to screens that would typically be able to attract more attention. Lastly, as for the manipulation of cognitive load, it might be interesting to test new support systems without adding a secondary task to the operator's daily operations during the tests as well. This way, one could first examine how an operator interacts with the new system in a familiar working environment. Afterward, it is still interesting to add the manipulation to see how operators would interact with the new system during more stressful periods by adding the secondary task to simulate a moment when the task demands would be very high.

5. Conclusion

Given that AI-based surveillance systems have not yet achieved 100% accuracy and it is unlikely that this will change in the near future, the current objective of these systems is to assist human operators. Therefore, it will remain important to investigate the influence of such systems on their human collaborator. The present study demonstrated how existing and future operator support systems in CCTV surveillance rooms can be tested using VR as a useful tool that can help to better design AI-driven support systems during the initial prototyping stages and to inform control room design guidelines. Additionally, by having investigated visual cueing techniques in the present study, the results provide insight into the possible advantageous effects of support systems in CCTV surveillance rooms, that is, an increase in the operators' monitoring task accuracy and a decrease in cognitive load measured by subjective reports and pupil size. As the results also demonstrate, future research should not merely evaluate the effectiveness of AI algorithms embedded in operator support systems, but also incorporate UX and usability research. This way, not only the effectiveness of the algorithm itself, but the full human-computer interaction and system design is investigated. Furthermore, the presented methodology can inspire research that focuses on different types of control rooms such as control rooms in nuclear power plants, in petro-chemical plants or even in air traffic control towers.

Acknowledgements

This study was funded by the imec.icon project SENSECITY (Project HBC.2018.0538, with support from VLAIO, the Flanders Innovation and Entrepreneurship agency).

We thank Floris De Smedt of Robovision for providing the algorithm and the resulting video sequence.

We thank our colleague Stephen Murphy, senior researcher at imec-mict-UGent for proofreading the manuscript.

Data availability

The data sets generated and analyzed during the current study are available in the OSF repository under project name *SenseCity Experiment 2* (see https://osf.io/ag3pf/).

References

Antonenko, P., Paas, F., Grabner, R., Van Gog, T., 2010. Using electroencephalography to measure cognitive load. Educational Psychology Review 22, 425–438.

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823 .

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F., 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neuroscience & Biobehavioral Reviews 44, 58–75.

Brouwer, A.M., Hogervorst, M.A., Van Erp, J.B., Heffelaar, T., Zimmerman, P.H., Oostenveld, R., 2012. Estimating workload using eeg spectral power and erps in the n-back task. Journal of neural engineering 9, 045008.

Chen, S., Epps, J., 2014. Using task-induced pupil diameter and blink rate to infer cognitive load. Human–Computer Interaction 29, 390–413.

Dadashi, N., Stedmon, A.W., Pridmore, T.P., 2013. Semi-automated cctv surveillance: The effects of system confidence, system accuracy and task complexity on operator vigilance, reliance and workload. Applied ergonomics 44, 730–738.

Dahlbäck, N., Jönsson, A., Ahrenberg, L., 1993. Wizard of oz studies—why and how. Knowledge-based systems 6, 258–266.

De Bruyne, J., Joundi, J., Morton, J., Van Kets, N., Van Wallendael, G., Talsma, D., Saldien, J., De Marez, L., Durnez, W., Bombeke, K., 2021. Smooth operator, in: International Conference on Human-Computer Interaction, Springer. pp. 233–240.

De Jong, P.J., Merckelbach, H., 1990. Eyeblink frequency, rehearsal activity, and sympathetic arousal. International Journal of Neuroscience 51, 89–94.

Di Nocera, F., Camilli, M., Terenzi, M., 2007. A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload. Journal of Cognitive Engineering and Decision Making 1, 271–285.

DiDomenico, A., Nussbaum, M.A., 2008. Interactive effects of physical and mental workload on subjective workload assessment. International journal of industrial ergonomics 38, 977–983.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A.R., Farhadian, M., Miyake, S., 2016. Analysis of the mental workload of city traffic control operators while monitoring traffic density: A field study. International Journal of Industrial Ergonomics 54, 170–177.

Gavas, R., Chatterjee, D., Sinha, A., 2017. Estimation of cognitive load based on the pupil size dilation, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE. pp. 1499–1504.

Grier, R.A., 2015. How high is high? a meta-analysis of nasa-tlx global workload scores, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, SAGE Publications Sage CA: Los Angeles, CA. pp. 1727–1731.

Grozdanovic, M., Janackovic, G.L., 2018. The framework for research of operators' functional suitability and efficiency in the control room. International Journal of Industrial Ergonomics 63, 65–74.

Hart, S.G., Staveland, L.E., 1988a. Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. North-Holland. volume 52 of *Advances in Psychology*, pp. 139–183. doi:https://doi.org/10.1016/S0166-4115(08)62386-9.

Hart, S.G., Staveland, L.E., 1988b. Development of nasa-tlx (task load index): Results of empirical and theoretical research, in: Advances in psychology. Elsevier. volume 52, pp. 139–183.

Hodgetts, H.M., Vachon, F., Chamberland, C., Tremblay, S., 2017. See no evil: Cognitive challenges of security surveillance and monitoring. Journal of applied research in memory and cognition 6, 230–243.

Hollis, M.E., 2019. Security or surveillance? examination of cctv camera usage in the 21st century.

Holm, A., Lukander, K., Korpela, J., Sallinen, M., Müller, K.M., 2009. Estimating brain load from the eeg. TheScientificWorldJOURNAL 9, 639–651.

Kenward, M.G., Roger, J.H., 1997. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics , 983–997.

Keval, H., Sasse, M., 2006. Man or gorilla? performance issues with cctv technology in security control rooms, in: 16th World Congress on Ergonomics Conference, International Ergonomics Association, Citeseer. pp. 10–14.

Klimesch, W., 1996. Memory processes, brain oscillations and eeg synchronization. International journal of psychophysiology 24, 61–100.

Krejtz, K., Duchowski, A.T., Niedzielska, A., Biele, C., Krejtz, I., 2018. Eye tracking cognitive load using pupil diameter and microsaccades with fixed gaze. PloS one 13, e0203629.

Ledger, H., 2013. The effect cognitive load has on eye blinking. The Plymouth Student Scientist 6, 206–223.

Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection, in: Proceedings of the IEEE international conference on computer vision, pp. 2980–2988.

Magliacano, A., Fiorenza, S., Estraneo, A., Trojano, L., 2020. Eye blink rate increases as a function of cognitive load during an auditory oddball paradigm. Neuroscience Letters 736, 135293.

Makowski, D., Pham, T., Lau, Z.J., Brammer, J.C., Lespinasse, F., Pham, H., Schölzel, C., Chen, S.A., 2021. Neurokit2: A python toolbox for neurophysiological signal processing. Behavior Research Methods , 1–8.

Noguchi, K., Gel, Y.R., Brunner, E., Konietschke, F., 2012. nparld: an r software package for the nonparametric analysis of longitudinal data in factorial experiments. Journal of Statistical software 50, 1–23.

Norris, C., McCahill, M., Wood, D., 2004. The growth of cctv: a global perspective on the international diffusion of video surveillance in publicly accessible space. Surveillance & Society 2.

Oberhauser, M., Dreyer, D., 2017. A virtual reality flight simulator for human factors engineering. Cognition, Technology & Work 19, 263–277.

Pelletier, S., Suss, J., Vachon, F., Tremblay, S., 2015. Atypical visual display for monitoring multiple cctv feeds, in: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1145–1150.

Pikaar, R., Lenior, D., Schreibers, K., de Bruijn, D., 2015. Human factors guidelines for cctv system design, in: Proceedings 19th Triennial Congress of the IEA, Melbourne, p. 9.

Puma, S., Matton, N., Paubel, P.V., Raufaste, É., El-Yagoubi, R., 2018. Using theta and alpha band power to assess cognitive workload in multitasking environments. International Journal of Psychophysiology 123, 111–120.

Recarte, M.Á., Pérez, E., Conchillo, Á., Nunes, L.M., 2008. Mental workload and visual impairment: Differences between pupil, blink, and subjective rating. The Spanish journal of psychology 11, 374–385.

Sauseng, P., Griesmayr, B., Freunberger, R., Klimesch, W., 2010. Control mechanisms in working memory: a possible function of eeg theta oscillations. Neuroscience & Biobehavioral Reviews 34, 1015–1022.

Setz, C., Arnrich, B., Schumm, J., La Marca, R., Tröster, G., Ehlert, U., 2009. Discriminating stress from cognitive load using a wearable eda device. IEEE Transactions on information technology in biomedicine 14, 410–417.

Sreenu, G., Durai, M.S., 2019. Intelligent video surveillance: a review through deep learning techniques for crowd analysis. Journal of Big Data 6, 1–27.

Stainer, M.J., Scott-Brown, K.C., Tatler, B.W., 2017. On the factors causing processing difficulty of multiple-scene displays. i-Perception 8, 2041669516689572.

Tauscher, J.P., Schottky, F.W., Grogorick, S., Bittner, P.M., Mustafa, M., Magnor, M., 2019. Immersive eeg: Evaluating electroencephalography in virtual reality, in: 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), IEEE. pp. 1794–1800.

Tremmel, C., Herff, C., Sato, T., Rechowicz, K., Yamani, Y., Krusienski, D.J., 2019. Estimating cognitive workload in an interactive virtual reality environment using eeg. Frontiers in human neuroscience 13, 401.

Van Acker, B.B., Parmentier, D.D., Vlerick, P., Saldien, J., 2018. Understanding mental workload: from a clarifying concept analysis toward an implementable framework. Cognition, technology & work 20, 351–365.

Vanneste, P., Raes, A., Morton, J., Bombeke, K., Van Acker, B.B., Larmuseau, C., Depaepe, F., Van den Noortgate, W., 2020. Towards measuring cognitive load through multimodal physiological data. Cognition, Technology & Work , 1–19.

Wanyan, X., Zhuang, D., Lin, Y., Xiao, X., Song, J.W., 2018. Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation. International Journal of Industrial Ergonomics 64, 1–7.

Warm, J.S., Parasuraman, R., Matthews, G., 2008. Vigilance requires hard mental work and is stressful. Human factors 50, 433–441.

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, 2005–2015.

Young, M.S., Brookhuis, K.A., Wickens, C.D., Hancock, P.A., 2015. State of science: mental workload in ergonomics. Ergonomics 58, 1–17.

Young, M.S., Stanton, N.A., 2001. Mental workload: theory, measurement, and application. International encyclopedia of ergonomics and human factors 1, 507–509.

Zheng, B., Jiang, X., Tien, G., Meneghetti, A., Panton, O.N.M., Atkins, M.S., 2012. Workload assessment of surgeons: correlation between nasa tlx and blinks. Surgical endoscopy 26, 2746–2750.

Zivkovic, Z., 2004. Improved adaptive gaussian mixture model for background subtraction, in: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., IEEE. pp. 28–31.