Touchscreens in Motion: Quantifying the Impact of Cognitive Load on Distracted Drivers

Xiyuan Shen*
Paul G. Allen School of Computer Science & Engineering | DUB Group
University of Washington
Seattle, Washington, USA
xyshen@uw.edu

Seokhyun Hwang*
The Information School | DUB Group
University of Washington
Seattle, Washington, USA
seokhyun@uw.edu

Junhan Kong
The Information School | DUB Group
University of Washington
Seattle, Washington, USA
junhank@uw.edu

Alexandre L. S. Filipowicz
Toyota Research Institute
Los Altos, California, USA
alex.filipowicz@tri.global

Andrew Best
Toyota Research Institute
Los Altos, California, USA
andrew.best@tri.global

Jean Costa
Toyota Research Institute
Los Altos, California, USA
jean.costa@tri.global

Scott Carter
Toyota Research Institute
Los Altos, California, USA
scott.carter@tri.global

James Fogarty
Paul G. Allen School of Computer Science & Engineering | DUB Group
University of Washington
Seattle, Washington, USA
jfogarty@cs.washington.edu

Jacob O. Wobbrock
The Information School | DUB Group
University of Washington
Seattle, Washington, USA
wobbrock@uw.edu

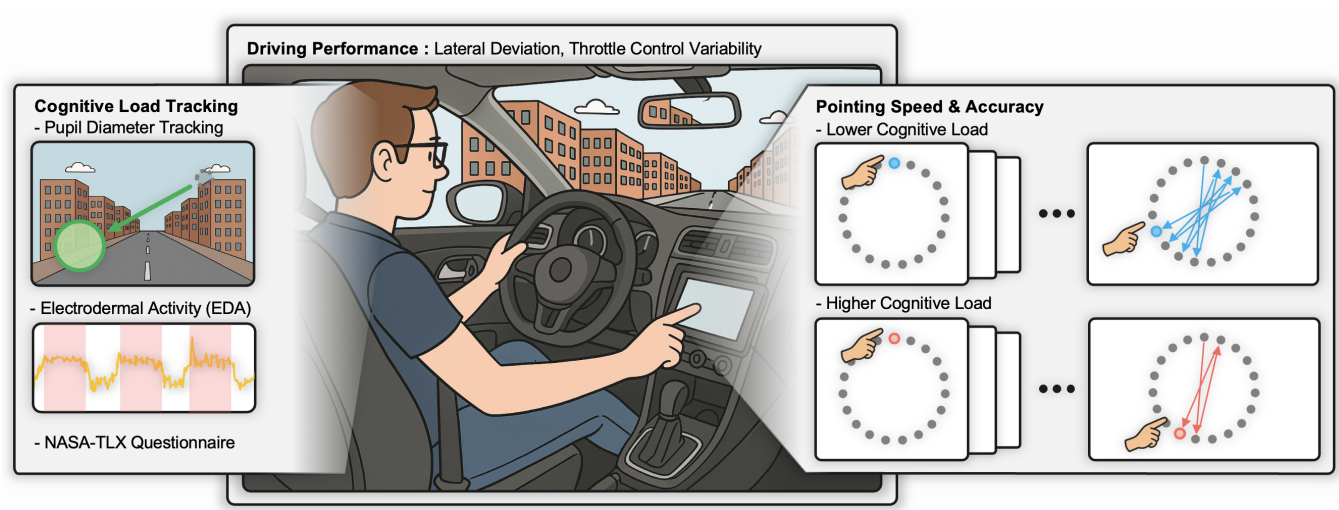

Figure 1: Conceptual diagram of the conducted study. It displays the three main methodologies used in this research. From left to right: pupil diameter, electrodermal activity (EDA), and the NASA-TLX questionnaire are shown as measures of cognitive load. Lateral deviation and throttle control variability are presented to assess driving performance, and the ISO 9241-9 "ring of circles" task used for Fitts' law analysis is included to quantify interaction with the touchscreen.

*These authors contributed equally to this work.

UIST '25, Busan, Republic of Korea

ACM ISBN 979-8-4007-2037-6/25/09
https://doi.org/10.1145/3746059.3747683

Abstract

This study investigates the interplay between a driver's cognitive load, touchscreen interactions, and driving performance. Using an N-back task to induce four levels of cognitive load, we measured physiological responses (pupil diameter, electrodermal activity), subjective workload (NASA-TLX), touchscreen performance (Fitts' law), and driving metrics (lateral deviation, throttle control). Our results reveal significant mutual performance degradation, with touchscreen pointing throughput decreasing by over 58.1% during driving conditions and lateral driving deviation increasing by 41.9% when touchscreen interactions were introduced. Under high cognitive load, participants demonstrated a 20.2% increase in pointing movement time, 16.6% decreased pointing throughput, and 26.3% reduced off-road glance durations. We identified a prevalent "hand-before-eye" phenomenon where ballistic hand movements frequently preceded visual attention shifts. These findings quantify the impact of cognitive load on multitasking performance and demonstrate how drivers adapt their visual attention and motor-visual coordination when cognitive resources are constrained.

CCS Concepts

• **Human-centered computing → Empirical studies in HCI**.

Keywords

In-vehicle touchscreen; cognitive load; Fitts' law; driver distraction; visual attention

ACM Reference Format:

Xiyuan Shen, Seokhyun Hwang, Junhan Kong, Alexandre L. S. Filipowicz, Andrew Best, Jean Costa, Scott Carter, James Fogarty, and Jacob O. Wobbrock. 2025. Touchscreens in Motion: Quantifying the Impact of Cognitive Load on Distracted Drivers. In *The 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25), September 28–October 01, 2025, Busan, Republic of Korea.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3746059.3747683

1 Introduction

Touchscreens have become increasingly prevalent in vehicles, providing access to various functions and information [4]. Despite their functionality, operating a touchscreen while driving introduces distractions that may negatively impact driver safety [57]. The National Highway Traffic Safety Administration (NHTSA) warns that visual distraction away from the road should be limited, recommending glances not exceed 2.0 seconds [74]. When secondary tasks like touchscreen interactions exceed driver processing capacity, information overload may occur [18, 107], impairing both driving safety and touchscreen efficiency. The result can be a frustrating user experience in the best case and a disastrous driving accident in the worst case.

Therefore, researchers have quantified factors affecting in-vehicle touchscreen interaction. Previous studies [3, 46, 55, 62, 109] have investigated how display size, location, and interface component design affect driving performance, focusing on isolated touchscreen tasks. However, there is growing recognition of the need to understand the specific mechanisms through which touchscreens and driving influence each other, necessitating an expansion of the research scope to include multitasking scenarios [28, 48, 75, 81]. In response, recent studies have explicitly examined user behaviors in multitasking contexts such as dialing or texting [9, 48, 58]. Although these studies explore some specific aspects of multitasking, there is a broader need to address a wider variety of secondary tasks.

To this end, we studied cognitive load and touchscreen performance to quantitatively assess various multitasking scenarios. Cognitive load quantifies the total mental resources engaged by a person across concurrent tasks and, in the context of driving, indicates a driver's information processing state and capabilities [7, 21, 113]. Our study aims to broaden the discussion on how multitasking-induced cognitive load influences driver behavior during touchscreen interactions by: (1) quantifying cognitive load, (2) quantifying driver touchscreen interaction performance, and (3) quantifying driving performance.

First, we manipulate different levels of cognitive load using an N-back task [78], a widely adopted paradigm for controlling working memory demands [31, 56, 92, 103, 104]. To measure the cognitive load, we collected physiological data from pupil diameter [44] and electrodermal activity (EDA) [88] and assessed participants' perceived mental workload using the NASA-TLX questionnaire [40, 41]. Second, to quantify drivers' touchscreen interaction performance, we used Fitts' law [30, 64] to model drivers' pointing performance, and used eye-tracking sensors to measure drivers' visual attention distribution between the road and the touchscreen. We specifically employed the ISO 9241-9 "ring of circles" task in which participants select targets arranged in a circular pattern [20, 47, 66, 91]. Third, we evaluated driving performance through two key metrics: steering and pedal control. Steering performance was assessed by lateral deviation from the road centerline, which measures how well the driver maintains lane position. Pedal performance was assessed through throttle control variability, which measures the consistency of gas pedal manipulation.

Our results show that touchscreen pointing throughput decreased by 58.1% during driving conditions (Section 4.3.1), and lateral driving deviation increased by 41.9% after introducing touchscreen interactions (Section 4.2.1). These results confirm that with limited information processing capacity, dual-task interplay degrades performance in both tasks.

Furthermore, high cognitive load significantly impacted drivers' touchscreen interaction. Drivers exhibited a 20.2% increase in pointing movement time and a 16.6% decrease in pointing throughput, from 2.42 bits/s to 2.01 bits/s (Section 4.3.2). Drivers' off-road glance duration decreased by 26.3%, from 1207 ms to 889 ms per gaze transition (Section 4.4), and induced a prevalent "hand-before-eye" phenomenon, where ballistic hand movements preceded visual attention shifts, intensifying to 71.9% under high cognitive load. These results emerged from an overloaded visual-motor system attempting to compensate for insufficient processing resources during complex multitasking conditions.

This paper provides the following research contributions:

- Quantitative empirical results showing mutual performance degradation between driving and touchscreen interactions.
- Quantitative empirical results showing how varying cognitive loads affect drivers' touchscreen pointing performance and visual attention.
- Mixed-methods results about how drivers adapt their visual attention to balance touchscreen interaction efficiency and driving safety under different cognitive loads.

Our findings highlight the potential of the hand-before-eye coordination pattern as a real-time cue for detecting elevated cognitive load during driving. We also propose design guidelines for interface strategies such as flattening interaction workflows to reduce multi-step tasks, adaptively accelerating target acquisition, and deploying load-sensitive alerts to interrupt prolonged off-road glances. Together, these implications support the design of future touchscreen interfaces that better balance usability and safety in high cognitive load contexts.

2 Related Work

Prior work related to this research primarily falls into three categories: (1) cognitive load when driving, (2) touchscreens in vehicles, and (3) Fitts' law in human-computer interaction. We take each of these in turn below.

2.1 Cognitive Load in Driving Contexts

Cognitive load has been a consistent topic for driving research [11], particularly within efforts to detect and reduce safety risks posed by distracted driving. Studies of driving in both real-world and simulated driving environments find that high cognitive load reduces people's attention to important cues in their surrounding environment and increases unsafe behaviors [5, 12, 23, 36, 39, 84]. For example, an on-road study by Harbluk et al. found that drivers performing a mental arithmetic task focus their gaze primarily on the center of the road. This shift in gaze led drivers to ignore peripheral cues indicated in their mirrors or at intersections and increased the frequency of dangerous hard braking events [39]. Moreover, a meta-analysis conducted by Caird et al. found that conversations while driving (i.e., on a cell phone or with a passenger) lower vigilance to external events, slow driver reaction times, and increase the chance of collisions [12].

In response to driving safety risks posed by cognitive load, regulators have proposed legislation mandating in-vehicle systems that detect and mitigate the effects of cognitive distractions [13]. However, distractions from cognitive load are challenging to detect because they represent an internal state rather than an overt external behavior (e.g., eyes off the road [33], hands off the wheel [73]) [37, 94]. As such, driving research has focused on identifying in-cabin methods to reliably measure and detect an individual's level of cognitive load.

There are three main methods used to measure cognitive load during driving [11, 60]: *self-report*, *task performance*, and *physiological measures*. *Self-report* methods, such as the Subjective Workload Assessment Technique (SWAT) [83], Workload Profile [99], Instantaneous Self-Assessment [96], and the NASA Task Load Index (NASA-TLX) [41, 79], provide *post hoc* assessments of subjective load but lack real-time sensitivity. *Task performance* measures examine changes in driving behavior (e.g., particularly situational awareness and reaction times to external events [5, 80]) and secondary-task performance (e.g., slower response times and reduced accuracy on peripheral detection or detection response tasks [8, 72]), reflecting impairments in cognitive control [24]. *Physiological measures* track sympathetic arousal (e.g., pupil diameter [31, 98], electrodermal activity [69, 70], heart rate [70]), eye-movement alterations [35, 36], and EEG spectral changes [16, 86]. Recent studies have combined these signals using machine learning to detect periods of elevated cognitive load (e.g., convolutional neural networks on eye-tracking videos [31], multimodal classifiers achieving up to 97% accuracy [42]).

Taken together, this previous work shows that cognitive load reduces a driver's attentional resources, limiting the bandwidth they have to act, such as performing unexpected evasive maneuvers or interacting with their car's touchscreen. It also highlights that measuring cognitive load requires a combined approach that includes subjective reports, behavior (particularly on secondary tasks), and physiology.

2.2 In-Vehicle Touchscreen Interfaces

Previous studies have investigated various factors influencing driver interaction performance with in-vehicle touchscreens. First, research has addressed the physical characteristics of touchscreens, such as their size and placement. Lamble et al. [55] reported that increased eccentricity of touchscreen locations from a driver's direct line of sight can lead to reductions in predicted collision time during tasks requiring sustained visual attention. Complementing this finding, Wittmann et al. [109] highlighted that the distraction caused by touchscreen interactions increases exponentially as the distance from the driver's primary visual field to the touchscreen interaction point grows. Furthermore, research by Ma et al. [62] provided evidence that, although larger screens (e.g., 10-inch or 17-inch displays) enhance the availability of information, they simultaneously exacerbate visual distractions compared to smaller screens (e.g., 7-inch or 9-inch).

Second, research has focused on the specific design attributes of touchscreen interfaces, including the arrangement and visual design of display elements. Studies by Nothdurft et al. [76] demonstrated that arranging interface elements closely together improves visual search efficiency; however, overly dense arrangements may impair the recognition of individual targets [102]. Additionally, Yoon et al. [111] found that visual features in vehicle instrument clusters (e.g., icon dimensions, density, and color variability) significantly affect the perceived visual complexity and consequently impact the efficiency of visual searches.

Although these studies provide valuable insights into the influence of touchscreen location, size, and visual design on driver performance, the majority of this research has predominantly concentrated on single-task scenarios [28]. Such scenarios effectively isolate driver interactions but fail to capture the multitasking complexity inherent in realistic driving situations. Addressing this gap, subsequent studies have begun to extend their scope to incorporate multitasking involving touchscreen interactions. Janssen et al. [48], for instance, explored drivers' cognitive chunk boundaries during dialing tasks, while Lee et al. [58] examined drivers' gaze patterns and task-switching behaviors during reading tasks performed while driving. These studies expanded the understanding of driver behavior in multitasking contexts, yet their findings remained confined to specific activities such as dialing or text reading, limiting their generalizability to realistic driving contexts.

2.3 Fitts' Law in Human-Computer Interaction

Fitts' law, originally proposed by Paul Fitts in 1954 [30], models the time (MT) it takes to perform rapid aimed movements to targets of a given width (W) at a given distance, *i.e.*, amplitude (A). Rapid aimed movements are those that can be guided by the actor and should be contrasted with ballistic movements, which are fully determined by their launch conditions. Widely adopted in human-computer interaction (HCI) [64], the "Shannon formulation" of Fitts' law [63] is expressed as:

$$MT = a + b \cdot \log_2 \left(\frac{A}{W} + 1 \right) \quad (1)$$

In Eq. 1, a and b are empirically derived regression coefficients, and the logarithmic term defines the nominal index of difficulty (ID) of the pointing task, measured in bits, with higher ID indicating a more challenging task. Although A, W, and ID specify the nominal task, actual performance often deviates: users may undershoot or overshoot the amplitude A, and may over- or under-use the target width W, possibly incurring errors. To account for these discrepancies, Crossman [14] proposed effective amplitude (A_e) and effective width (W_e), subsequently validated in prior work [64, 65, 91, 105]. In each $A \times W$ condition, A_e is the mean of actual movement distances, and W_e is computed based on the standard deviation of endpoints σ around target centers, and amounts to $\sigma\sqrt{2\pi e}$, a constant related to the entropy of a standard normal distribution [110]. This correction ensures W_e reflects true pointing precision rather than nominal target width. By computing ID_e, researchers integrate users' speed–accuracy biases into a unified throughput metric [65, 91, 112]:

$$ID_e = \log_2 \left(\frac{A_e}{W_e} + 1 \right) \quad (2)$$

To minimize directional biases and enhance reliability, the ISO 9241-9 standard recommends the "ring of circles" paradigm, which arranges targets evenly spaced along the circumference of a circle [20, 47, 66, 67, 91]. Participants begin from the top target and sequentially move across the diameter of the circle, tapping each target around the ring in a clockwise fashion. This design mitigates systematic directional biases that could influence movement time, as it requires movements in multiple directions.

3 Experiment Method

This section describes the methodology of the experiments, including participants (Section 3.1), apparatus and sensors (Section 3.2), experimental design and data processing (Section 3.3 — 3.4), procedure (Section 3.5), and statistical methods (Section 3.6).

3.1 Participants

Sixteen participants (10 female, 6 male) with a mean age of 25.8 years (SD = 4.1) and an average driving experience of 5.7 years (SD = 3.0) were recruited. Driving frequency was distributed as: daily (4), two or three times weekly (3), once a week (3), monthly (3), and less than monthly (3). All participants possessed valid driver's licenses. Participants with myopia were instructed to wear contact lenses. All participants could terminate the experiment at any time and received a $40 gift card as compensation. The study was approved by our university's Institutional Review Board.

3.2 Apparatus

We developed a driving simulation integrated with eye-tracking, electrodermal activity (EDA) sensors, and vision-based finger tracking sensors, along with an auditory N-back task and a touchscreen-based Fitts' law task (Figure 2).

3.2.1 Simulation Environment. The driving simulation was developed using the open-source driving simulator *CARLA* [19]. For this study, we utilized *CARLA*'s *Town10* map, which features an urban setting, as illustrated in Figure 2. Participants followed a predefined continuous route along the outer perimeter road. Participants used the *Fanatec Podium Wheel Base DD2* steering system and the *Fanatec Clubsport Pedals V3*, mounted on a *Trak Racer TR80* simulator frame, as shown in Figure 2. A 38-inch widescreen monitor was also attached to the TR80 frame, with the seat position, steering wheel height, and pedal distance adjustable to accommodate individual participant preferences.

3.2.2 Sensor Setup. To measure drivers' physiological states and touchscreen interactions, we integrated three sensors. Eye-gaze data were captured using the *Tobii Pro Glasses 2*, which tracks eye movements to determine a driver's visual attention. It records gaze data at 100 Hz and captures 1920×1080 video at 25 fps from the user's viewpoint.

Cognitive load was quantitatively assessed using pupil diameter and EDA [10, 89, 100]. Pupil diameter was measured using the *Tobii Pro Glasses 2*, while EDA data were recorded at 15 Hz using the *EmotiBit* sensor [71], an open-source wearable device secured to participants' wrists for physiological data collection.

To capture finger movement distance and timing for the Fitts' law analysis, we used a *ZED 2i* stereo camera combined with Google's *MediaPipe* hand-tracking algorithm. The *ZED 2i* offers a 110°×70°×120° field of view, records binocular images at 2560×720 resolution and 60 fps, and provides 16-bit depth data processed via the *ZED SDK* and *MediaPipe* to track index finger trajectories in 3-D space.

3.2.3 N-back Task. The N-back task for this study required auditory responses from participants. A hi-fi speaker was integrated into the simulation setup to deliver audible stimuli to participants in the form of sequential random numbers, which were generated and converted into auditory form using *Google Text-to-Speech*, with each number presented at intervals of 2.5 seconds. Participants verbally responded with the number presented N numbers behind the current number, and their responses were manually recorded by the experimenter.

3.2.4 Fitts' Law Task. We implemented the "ring of circles" ISO 9241-9 task [20, 47, 91]. A *Samsung Galaxy Tab S7 FE 12.4"* was positioned to the participant's right to enable comfortable touchscreen interaction, as illustrated in Figure 2. This tablet features a 12.4-inch display with a 16:10 aspect ratio and a resolution of 2560×1600 pixels. Active targets for selection appeared in Dodger blue, while inactive targets were displayed in gray. Incorrect selections caused the target to briefly flash red (for 100 ms), accompanied by an error tone, providing immediate feedback to the participant regarding misses.

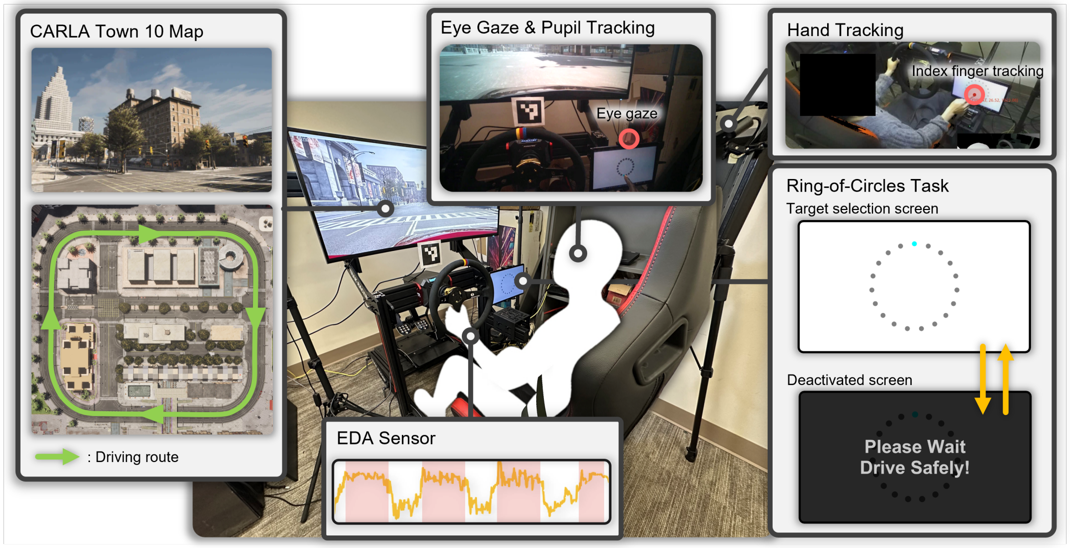

Figure 2: Driving simulator and sensor system implemented for the experiment. Participants drove along a predefined route in an urban environment, shown on the left side of the figure. Simultaneously, they were instructed to perform the ring of circles task using a touchscreen located to the right of the steering wheel. During this process, a sensor for measuring electrodermal activity was attached to the participant's wrist, a depth camera for tracking hand movements was installed above, and eyeglass-mounted sensors were worn to track gaze and measure pupil diameter.

Baseline Non-Driving Target Selection Task. To establish a baseline for target selection performance without driving, we conducted a small Fitts' law experiment with eight participants. Participants performed the same Fitts' law task using identical experimental apparatus but without simultaneous driving or N-back tasks. The baseline experiment yielded an average throughput of 5.78 bits/s (SD = 0.61), which is in keeping with prior work on touchscreen target selection [6, 49, 87]. This result provided us with a baseline to quantify the impact of driving and cognitive load on touchscreen interaction.

3.2.5 Questionnaire. Participants completed three questionnaires after each cognitive load condition:

- **NASA-TLX** [40, 41]: Measures subjective workload under each condition.
- **Short Stress State Questionnaire (SSSQ)** [43]: Measures participants' perceived stress levels.
- **International Positive and Negative Affect Schedule Short Form (I-PANAS-SF)** [50]: Assesses emotional states and affective responses.

3.3 Experiment Design

Our study aimed to observe and analyze the behaviors of drivers interacting with touchscreens under varying cognitive load conditions. We systematically induced four distinct cognitive load conditions:

- **No N-back Task**: Participants performed driving and touchscreen target selections *without* the N-back task.
- **0-Back**: Participants repeated the number they *just* heard while driving and performing touchscreen target selections.
- **1-Back**: Participants responded with the number presented *one* number earlier while driving and performing touchscreen target selections.
- **2-Back**: Participants responded with the number presented *two* numbers earlier while driving and performing touchscreen target selections.

Each participant experienced all cognitive load conditions once, with the sequence randomized using a balanced Latin square.

Under each cognitive load condition, participants completed the Fitts' law task using the "ring of circles" setup. Two movement amplitudes were used: $A_1 = 70$ mm and $A_2 = 120$ mm. Each amplitude was combined with three target widths: $W_1 = 6.9$ mm, $W_2 = 8.6$ mm, and $W_3 = 10.3$ mm. This design resulted in six distinct $A \times W$ combinations, representing nominal indices of difficulty (IDs) ranging from 2.96 bits to 4.20 bits. As shown in Figure 2, 19 targets were arranged for each $A \times W$ set, with the first three targets used as practice for participant warm-up, and the remaining 16 targets used for data collection. Excluding practice trials, the total number of trials was (16 participants) × (4 cognitive load conditions) × (6 $A \times W$ sets) × (16 targets) = 6,144.

3.4 Data Collection and Analysis

This section presents the data collected under each experimental condition (Section 3.3) and the corresponding processing methods.

3.4.1 Physiological Signals. We utilized pupil diameter and EDA to verify changes in participants' cognitive load. A preliminary user study supporting this approach's feasibility is included in the Appendix. Standard processing pipelines [52] were employed to process pupil diameter data. Blinks were removed and linearly

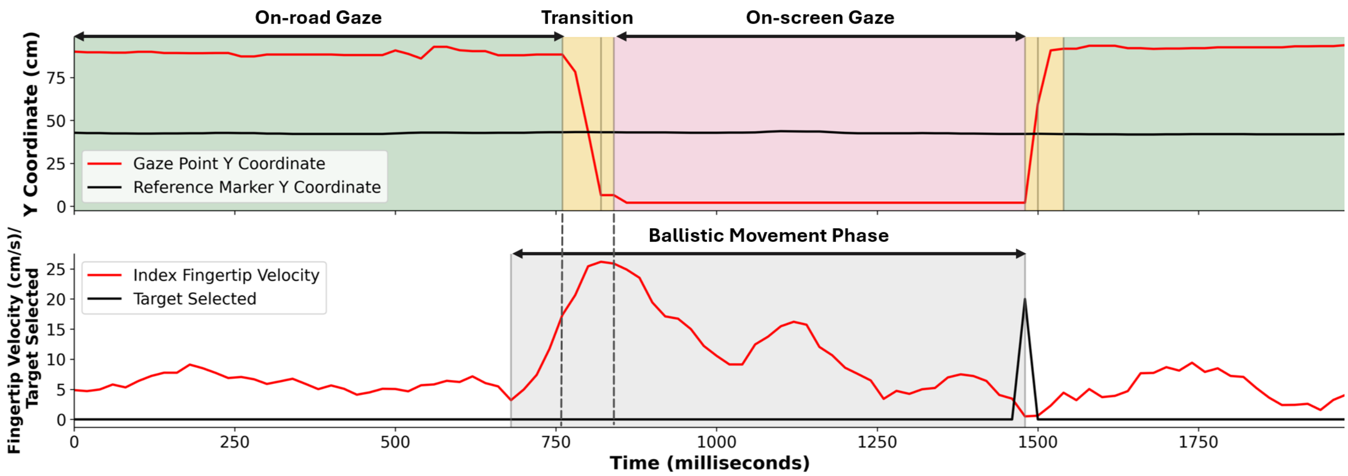

Figure 3: Gaze focus transitions and finger ballistic movement phases during pointing operations. The visual focus of attention is determined by the vertical relative position between the gaze point and the reference marker. The ballistic movement phase begins with the most dominant velocity impulse of the fingertip and concludes when the target is selected on the touchscreen. The dotted lines indicate that ballistic movements begin before the visual focus transitions to the screen, even before the transition phase. We discuss this finding in our Discussion.

interpolated. The pupil data was low-pass filtered at 10 Hz and Z-scored within each participant. To assess cognitive load-induced changes, we subtracted a baseline period (two seconds before the start of each block) from pupil diameter measurements.

3.4.2 Driving Performance. We quantified driving performance using two metrics: lateral deviation from the road centerline and throttle control variability. Lateral deviation measures how well the driver maintains lane position. The throttle control variability evaluates how consistently the user manipulates the pedal. Both throttle and brake inputs were normalized to the $[0, 1]$ range, with 0 indicating no pressure and 1 indicating full depression.

Lateral deviation was calculated as the frame-by-frame distance between the vehicle center and the road centerline. Curved segments were excluded to eliminate the confounding effects of their inherent cognitive demands, as noted by previous research [77]. For each condition, mean deviation was computed and normalized by subtracting the participant's baseline, recorded after the practice session without any secondary task. Throttle control variability was evaluated using the standard deviation of throttle input, also excluding curved segments. These values were similarly normalized using each participant's baseline driving data.

3.4.3 Fitts' Law Task Metrics. We employed hand tracking to calculate the actual moving distance for each pointing action and averaged these measurements to determine the effective amplitude. Unlike traditional Fitts' law tasks [64] where participants perform a series of continuous pointing actions during a block, our experimental design required participants to alternate between steering wheel control and pointing tasks. Consequently, while the origin of pointer movement during continuous target selection would typically be the previous target circle on the "ring of circles," a substantial proportion of trials—those originating from the steering wheel to the touchscreen—had effective amplitudes (A_e) much greater than others—those originating from the touchscreen itself.

Research on feedback control of hand movements [15] reveals that an aimed pointing movement consists of a series of discrete corrective motions decreasing in magnitude, called "submovements." By identifying the starting point of the initial ballistic phase of the movement, we determined the actual starting position and movement time for each pointing trial, enabling us to calculate A_e, even for trials originating from the steering wheel. Thus, from a given $A \times W$ ring-of-circles condition, two $(ID_e, \overline{MT})$ ordered pairs could be generated, one for trials originating from the touchscreen, and another for trials originating from the steering wheel.

First, we calculated the movement distance for each pointing trial. Using RGB images captured by a *ZED 2i* camera processed through *Mediapipe* [61], we obtained the 2-D position (x_{2d}, y_{2d}) of the finger in each frame. Combined with camera intrinsic parameters (f_x, f_y, c_x, c_y) and synchronized depth data from *ZED 2i* camera, we derived the real-time 3D position of the index finger according to:

$$X = \frac{(x_{2d} - c_x)Z}{f_x} \tag{3}$$

$$Y = \frac{(y_{2d} - c_y)Z}{f_y} \tag{4}$$

$$Z = D(x_{2d}, y_{2d}) \tag{5}$$

where $D(x_{2d}, y_{2d})$ represents the depth value at the 2-D finger coordinates.

As shown in Figure 3, by calculating the 3-D motion velocity of the fingertip, we employed the *find_peaks* function from Python's SciPy library [101] to detect the starting point of the dominant impulse in the ballistic phase preceding each touch action. This point was established as the origin of finger movement for each trial. The distance from this origin to the target circle on the touchscreen represented the actual movement distance for each pointing action, providing a more accurate measurement than screen-based calculations alone. A similar approach to isolating aimed pointing movements from continuous motion was used in the *Input Observer* by Evans and Wobbrock [26].

Next, we applied a Gaussian Mixture Model [85] for binary classification of all pointing trials per participant at each amplitude level. This yielded clear movement distance classification thresholds to differentiate the origin of each pointing action as either from the touchscreen or from the steering wheel. For each $A \times W$ ring-of-circles condition, if five or more trials of one origin type were produced, we used all trials of that type to calculate A_e, W_e, ID_e, and $\overline{MT}$. A_e averaged the movement distance, while W_e was calculated as $\sqrt{2\pi e} \times SD_{x,y}$, where $SD_{x,y}$ is the bivariate standard deviation of touch positions in the x and y directions on the touchscreen [110]. Blocks with fewer than five trials for a particular origin type were discarded as outliers. For each participant under each cognitive load, this ultimately provided 6-12 data points (1-2 for each $A \times W$ ring-of-circles condition) to fit a Fitts' law model.

3.4.4 Gaze and Focus of Attention. Gaze and eye movement serve the purpose of identifying an individual's focus of attention (FoA) [93]. In our study, FoA referred to the area where the gaze was concentrated while participants simultaneously drove and interacted with a touchscreen.

We defined "visual distraction" in relation to the primary task of driving, measured as any period when the driver's gaze was directed away from the road. Although both driving and touchscreen interactions were given as experimental tasks, we considered the gaze directed at the touchscreen as a "visual distraction" from a driving safety perspective. We classified drivers' gaze point (FoA) into three categories:

- **On-road gaze**: The FoA was on the road.
- **On-screen gaze**: The FoA was on the touchscreen.
- **Gaze transition**: The FoA was shifting between driving and touchscreen pointing tasks.

We positioned an *ArUco* marker [34] at the bottom edge of the driving simulator monitor, with the touchscreen located at the lower right of the monitor. As shown in Figure 3, we compared the relative position of the participant's gaze point to the *ArUco* marker in real-time using the first-person camera data from *Tobii Pro Glasses 2* to determine whether the user's FoA was on-road (gaze point higher than the *ArUco* marker) or on-screen (gaze point lower than the *ArUco* marker). Additionally, we calculated gaze point velocity and identified rapid eye movements exceeding three standard deviations during FoA transitions as the gaze transition period. We defined each "visual distraction duration" as the amount of time participants' gaze points left the road on the driving simulator monitor, which included a period of on-screen gaze and two rapid gaze transition periods.

3.5 Procedure

Prior to the experiment, participants received detailed explanations of the experimental procedures and data collection practices and provided informed consent. They were then equipped with an *EmotiBit* sensor and *Tobii Pro Glasses 2*. Participants adjusted the seat, steering wheel, and pedals to their preferred positions.

Participants were instructed to prioritize driving safely in all tasks by maintaining a stable speed, staying within their lane, and avoiding abrupt accelerations or decelerations. Performance on the N-back and pointing tasks began simultaneously with the driving simulation. Participants were encouraged to respond accurately to the N-back task and to perform touchscreen interactions for the target selection task "quickly and accurately." After completing a single ring-of-circles for one nominal ID, participants were given a 30-second rest before proceeding to the next set. During each rest period, participants continued driving without performing the N-back or touchscreen tasks, and the touchscreen interface was deactivated (see Figure 2).

Before the study began, participants completed three practice blocks, each lasting a minimum of five minutes, to familiarize themselves with the experimental setup and tasks. These included: (1) practice with the N-back task, (2) driving practice along the experimental route (Figure 2), and (3) practice driving with the touchscreen target selection task. Participants could request additional practice if needed, and five of them did so. After the practice sessions, baseline driving behavior was recorded by having participants drive the specified route without additional tasks. Participants then sequentially experienced each cognitive load condition while using the touchscreen, completing the three questionnaires after each. Sufficient rest periods were ensured between conditions to mitigate fatigue.

3.6 Statistical Analysis

Our statistical analysis consists of two primary components. We first examined whether driving and touchscreen interaction mutually influence each other. We then assessed how cognitive load levels impact driver performance across multiple measures.

3.6.1 Driving With vs. Without Touchscreen Interaction. We investigated the effect of touchscreen interaction on driving performance without the N-back task. Using driving without touchscreen interaction and driving with touchscreen interaction as independent variables, we examined their impact on driving performance metrics. These include driving deviation and throttle control variability. A within-subjects design with two conditions was employed for this analysis.

3.6.2 Touchscreen Interaction With vs. Without Driving. We compared touchscreen pointing performances when driving versus when not driving, both without the N-back task. This analysis examined the effects of driving on pointing movement time, error rate, and throughput. A between-subjects design with two conditions was employed.

3.6.3 Cognitive Load Impact. To assess the impact of cognitive load on drivers' performance across multiple tasks, we employed a within-subjects design with repeated measures across four cognitive load conditions. In this experiment, participants completed 6,144 trials in total, excluding practice trials.

The independent variable was cognitive load level, manipulated through four conditions: driving without an N-back task, with a 0-back task, with a 1-back task, and with a 2-back task. Dependent variables comprised four categories: (1) cognitive load verification measures (NASA-TLX ratings, pupil diameter, and EDA); (2) driving performance metrics (lateral driving deviation and throttle control variability); (3) touchscreen interaction performance (movement time, error rate, and pointing throughput); and (4) visual attention allocation (proportional distribution across FoAs, visual distraction duration, and the proportion of prolonged visual distraction).

Our EDA sensor malfunctioned for one participant under one N-back condition. Three users experienced a loss of *ArUco* marker detection in all of their first-person videos from the *Tobii Pro Glasses 2*. We removed the affected trials accordingly.

3.6.4 Statistical Methods. We used the Shapiro-Wilk test [90] to verify whether model residuals violated normality. The residuals of NASA-TLX ratings, throttle control variability, and the proportion of prolonged visual distraction were non-normally distributed, while other dependent variables did not violate normality.

For normal data, we used linear mixed models (LMM) [106] for within-subjects designs and linear models (LM) [27] for between-subject designs. We used Type III F-tests [22] to assess statistical significance across conditions, followed by *post hoc* pairwise comparisons.

For non-normal data, we employed the Friedman test [32] to examine significance across cognitive load levels, followed by Wilcoxon signed-rank tests [108] for *post hoc* pairwise comparisons.

All *post hoc* tests were protected against Type I errors using Holm's sequential Bonferroni procedure [45]. We report means and standard deviations (SD) for normally distributed variables and medians and interquartile ranges (IQR) for non-normally distributed variables.

4 Results

In this section, we present our results, categorized by cognitive load, driving performance, touchscreen performance, and focus of attention. Together, these results paint a picture of how cognitive load affects driving and interacting with a touchscreen, and how the latter two affect each other.

4.1 Cognitive Load

We analyzed participants' ratings on the NASA-TLX Likert items, changes in pupil diameter, and electrodermal activity (EDA) under different cognitive load conditions to verify that we successfully manipulated cognitive load. Recall that in our four N-back conditions, we required participants to complete driving and touchscreen interaction with no N-back task, with a 0-back task, with a 1-back task, and with a 2-back task.

4.1.1 NASA-TLX Mental Load. Mental load scores were particularly high when performing the 2-back task, with the median reaching 16 (IQR = [14.0,17.0]) out of 20. The 1-back task had a median score of 12.5 (IQR = [8.0,14.25]), followed by the 0-back task at 10 (IQR = [6.75,12.0]) and the no N-back task setting at 9.5 (IQR = [7.75,11.0]). There was a statistically significant difference in mental load across cognitive load conditions ($\chi^2(3,N{=}16) = 27.27$, $p < .001$), as shown in Figure 4(a). This result reflects participants' subjective experience that different N-back tasks imposed different degrees of cognitive load. *Post hoc* pairwise comparisons indicated significant mental load differences between all pairs of N-back tasks ($p < .01$), except between the 0-back task and without any N-back task.

4.1.2 Pupil Diameter and Electrodermal Activity. There were significant differences in pupil diameter ($F(3,45) = 2.68$, $p < .05$) and EDA ($F(3,45) = 3.47$, $p < .05$) across cognitive load conditions, as is shown in Figures 4(b) - (d). These physiological sensor signals provide objective evidence that the different cognitive load conditions did indeed impose varying levels of cognitive load. *Post hoc* pairwise comparisons showed similar results as the self-reported mental load measures. For example, the average pupil diameter was significantly higher during the 2-back task condition ($p < .05$) compared to the baseline no N-back task condition.

The results justify our approach of treating "different cognitive load levels" as our independent variable rather than simply "different N-back task conditions."

4.2 Driving Performance

We first examined how touchscreen interaction affects driving performance by comparing driving metrics with and without touchscreen interaction. Subsequently, we investigated driving performance under different cognitive load levels.

4.2.1 Touchscreen Effects on Driving Performance.

Lateral Deviation from the Road Centerline. Participants demonstrated 12.6% lower lateral deviation when driving without touchscreen interaction ($M = 0.355m$, $SD = 0.086m$) than with touchscreen interaction ($M = 0.406m$, $SD = 0.122m$). Statistical analysis revealed a significant difference ($F(1,15)=5.22$, $p < .05$), indicating that touchscreen interaction resulted in greater lateral deviations from the road centerline. This finding suggests that engaging with touchscreen interfaces impairs lateral vehicle control, despite explicit instructions for participants to prioritize driving safety as their primary task.

Throttle Control Variability. Analysis of throttle control variability showed comparable median deviations between the driving with touchscreen interaction condition ($Med = 0.052$, $IQR = 0.024$) and that without touchscreen interaction ($Med = 0.053$, $IQR = 0.076$). A Wilcoxon signed-rank test indicated no significant difference in throttle control variability between driving with a touchscreen and driving without a touchscreen.

4.2.2 Cognitive Load Effects on Driving Performance.

Lateral Deviation from the Road Centerline. There was minimal difference in lateral deviation across different cognitive load conditions. Results showed no statistically significant effects of cognitive load on centerline deviation. These results suggest that driving precision in terms of lateral control was not substantially compromised by cognitive load.

Throttle Control Variability. The Friedman test indicated no statistically significant effects of cognitive load on throttle control variability. These findings parallel our results on lateral deviation, demonstrating that drivers' ability to operate the pedal was not significantly impaired by increasing cognitive loads.

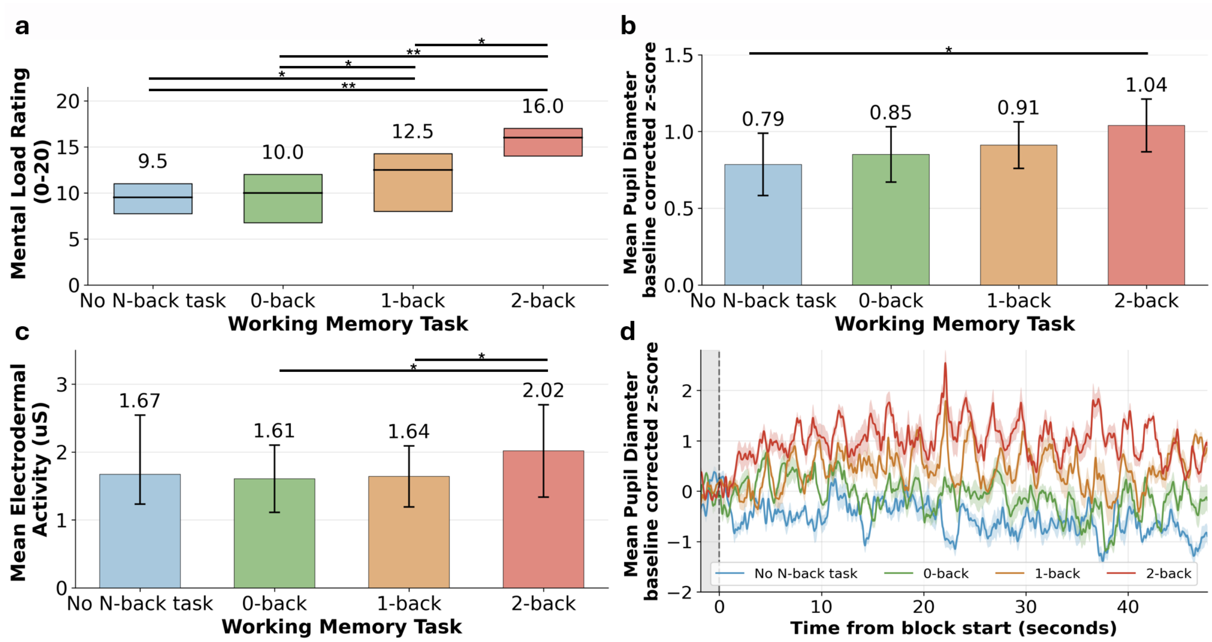

Figure 4: Subjective NASA-TLX ratings of cognitive load levels and physiological measurements across different *N*-back conditions. (a) Mental workload. Boxes represent quartiles. (b) Mean pupil diameter. Error bars represent standard deviation. (c) Mean electrodermal activity (EDA). (d) Mean changes in pupil diameter during Fitts' law Task blocks under different *N*-back task conditions, with shading indicating the standard error. ***:$p < .001$,**:$p < .01$,*:$p < .05$.

4.3 Touchscreen Pointing Speed and Accuracy

We first compared pointing performance under driving scenarios with the baseline throughput obtained by seated but non-driving participants (see Section 3.2.4). We then further analyzed the movement time, accuracy, and throughput of participants' touchscreen pointing while driving under different cognitive loads.

4.3.1 *Effects of Driving on Touchscreen Pointing.*

Speed and Accuracy. The average movement time rapidly increased from 564 ms (SD = 92) without driving to 1140 ms (SD = 205) with driving. Movement time was statistically significantly slower under driving conditions ($F(1,22) = 56.1$, $p < .001$). Similarly, error rate increased from 5.6% (SD = 4.0) without driving to 11.6% (SD = 9.2) with driving. However, due to substantial individual differences in target selection precision, this difference was not statistically significant.

Throughput. Without driving, the average throughput was 5.78 bits/s (SD = 0.66), while driving without an *N*-back task reduced target selection throughput by 58.1% to 2.42 bits/s (SD = 0.52). Statistical analysis revealed that this was a significant reduction in pointing throughput ($F(1,22)$=187.8, $p < .001$).

4.3.2 *Effects of Cognitive Load on Touchscreen Pointing.*

Speed and Accuracy. There was a significant effect of cognitive load on movement time ($F(3,N{=}45) = 10.13$, $p < .001$), as shown in Figure 5. *Post hoc* pairwise comparisons showed during the 2-back task, the movement time was significantly longer than that without

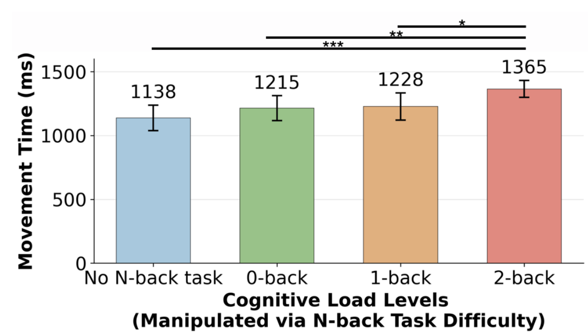

Figure 5: Movement time. Error bars represent standard deviation. ***:$p < .001$,**:$p < .01$,*:$p < .05$.

an *N*-back task ($t(45) = 5.41$, $p < .001$), with a 0-back task ($t(45)$ = -3.592, $p < .005$), and with a 1-back task ($t(45)$ = -3.284, $p < .01$). As the cognitive load increased, the mean of 1140 ms (SD = 205) for movement time without an *N*-back task significantly increased by 20.2% to 1370 ms (SD = 137) for pointing while performing the 2-back task.

Although increasing cognitive load resulted in increased movement time, error rates showed no significant difference across different cognitive load levels. The error rate during the 2-back task was 12.2% (SD = 9.3%), similar to the 11.6% (SD = 9.2%) under the no *N*-back task condition. The large standard deviations indicate substantial individual differences in pointing accuracy.

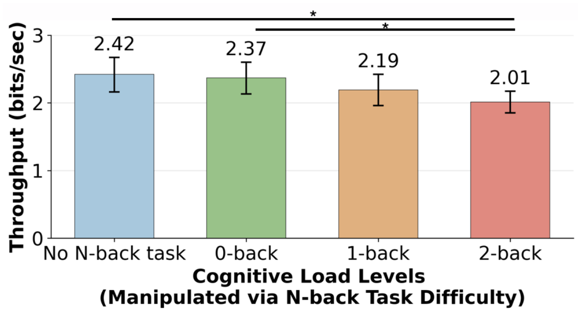

Figure 6: Pointing Throughput. Error bars represent standard deviation. ***:**$p < .05$.

These results suggest that under higher cognitive loads, participants' motor systems tended to sacrifice pointing speed while maintaining pointing accuracy.

Touchscreen Pointing Throughput. A main motivation for this work was to investigate whether different cognitive loads during driving influence throughput, the combined speed-accuracy measure of pointing efficiency produced by Fitts' law. Figure 6 shows the effect of cognitive load on throughput.

Without an N-back task, throughput was 2.42 bits/s (SD = 0.52). Throughput for the 0-back task was similar (M = 2.37, SD = 0.48), followed by the 1-back task (M = 2.19, SD = 0.47) and 2-back task (M = 2.01, SD = 0.32). There was a significant effect of cognitive load on throughput (F(3,45) = 4.10, $p < .05$). *Post hoc* pairwise comparisons showed that when the cognitive load was highest (*i.e.*, during the 2-back task), throughput was significantly lower than without an N-back task ($t(45)$ = 3.15, $p < .05$), and with a 0-back task ($t(45)$ = 2.76, $p < .05$). The standard deviations indicate substantial individual differences in pointing performance. However, the statistically significant results demonstrate that performance degradation consistently occurs with increased cognitive load.

4.4 Focus of Attention

When participants were driving while pointing on the touchscreen, their visual attention switched repeatedly between the road and the touchscreen. Based on eye tracking results, we calculated the distribution of participants' visual attention, the duration of each visual distraction, and the proportion of prolonged visual distractions. We checked if these metrics had significant associations with cognitive load.

First, we divided foci of attention into on-road gaze, on-screen gaze, and gaze transitions, as shown in Figure 7. There was no significant difference in the proportional duration of the three foci of attention under different cognitive loads. However, as cognitive load increased, the proportion of participants' visual attention on the road increased from 42.8% to 49.8%, with corresponding decreases in the proportion of gaze time on-screen and in transition.

One possible reason to explain why participants spent a smaller proportion of time on pointing tasks is that they reduced their frequency of switching from driving to touchscreen pointing. Another possibility is that high cognitive load caused participants to

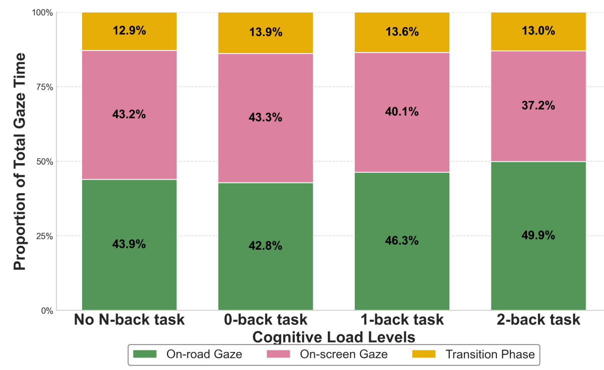

Figure 7: Proportion of focus of visual attention across different N-back conditions.

shorten the duration of each focus of attention on the touchscreen and gaze transition.

Significance testing showed no difference in the frequency of gaze switches from driving to the screen across different cognitive loads. Switch frequency ranged from 24.5 times per minute to 27.3 times per minute, with no consistent trend of change as cognitive load levels increased. Thus, the following paragraphs quantitatively check the visual distraction duration.

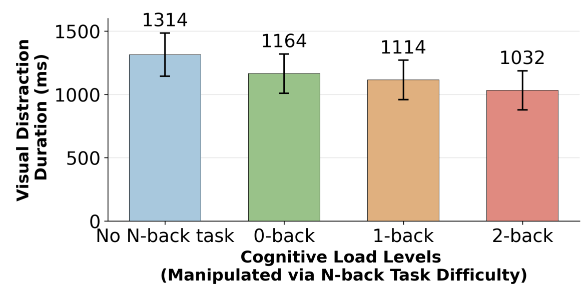

Figure 8: Mean duration of each visual distraction across different N-back conditions.

4.4.1 Visual Distraction Duration. As defined earlier, visual distraction refers to periods in which the driver's gaze is directed away from the road toward the touchscreen. As is shown in Figure 8, there is a significant difference between visual distraction durations under different cognitive loads (F(3,36) = 12.98, $p < .05$). When cognitive load increased, participants' gaze remained on the touchscreen for shorter periods before they needed to return their visual attention to the driving task. $Post - hoc$ pairwise comparisons showed that with a 2-back task, the mean visual distraction duration (M = 889 ms, SD = 327) was significantly shorter ($t(36)$ = 3.52, $p < .01$) than without a N-back task (M = 1207 ms, SD = 347).

4.4.2 Prolonged Visual Distraction. N-back task$N$-back task The NHTSA proposed a 2-second rule through a series of quantitative studies: each visual distraction duration for drivers should be less than 2 seconds to ensure safe driving [74]. Therefore, we calculated

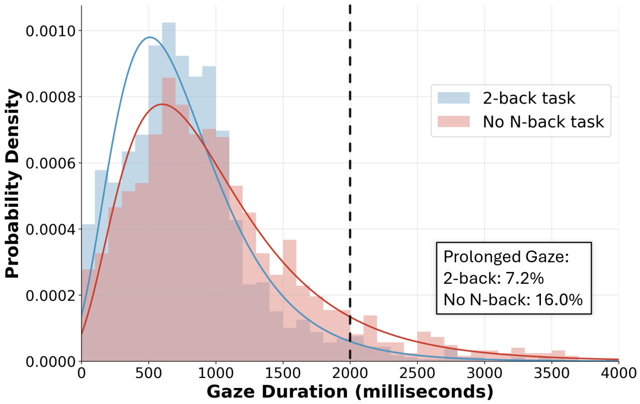

Figure 9: Distribution of visual distraction duration from driving under no *N*-back task and with 2-back task conditions, with curves representing lognormal distribution fits. Under the no *N*-back task condition, prolonged visual distractions occur significantly more frequently.

the proportion of participants' visual distractions exceeding 2000 ms to check whether our participants adhered to the 2-second rule.

There is a significant difference in the proportion of prolonged visual distractions (defined as greater than 2000 ms) under different cognitive loads ($\chi^2(3{,}N{=}13) = 9.09$, $p < .05$), shown in Figure 9. Without a N-back task, 16.0% (SD = 13.3%) of visual distractions exceeded 2 seconds, which is defined as potentially hazardous [74]. However, with a 1-back or 2-back task, the proportion of prolonged visual distractions decreased to 9.4% (SD = 9.5) and 7.2% (SD = 6.8), respectively. *Post hoc* pairwise comparisons revealed that the proportion of prolonged visual distractions was significantly higher without the N-back task compared to the condition with a 2-back task ($t(36) = -3.385$, $p < .05$). High cognitive load reduced visual distraction duration, but given that participants maintained driving safety to the same degree under different cognitive loads, a lower rate of prolonged visual distraction clearly cannot indicate safer driving.

4.5 Hand-Before-Eye Coordination Patterns

When participants were driving while performing a pointing task on the touchscreen, their visual attention switched repeatedly between the road and the touchscreen, while simultaneously executing target selection hand movements. We analyzed the temporal sequence between each ballistic hand movement and visual transition to the touchscreen to examine coordination between participant hand movements and visual attention allocation, as illustrated in Figure 3.

Unexpectedly, even without an N-back task, ballistic movements preceded gaze transitions to the touchscreen in 62.9% (SD = 39.6) of the pointing trials. This "hand-before-eye" pattern remained consistent across different cognitive load conditions: 64.4% (SD = 35.8) with a 0-back task, 71.9% (SD = 27.2) with a 1-back task, and 71.1% (SD = 24.6) with a 2-back task. Although there were no statistically significant differences between conditions, prevalence of this pattern was consistently high and showed a gradual increase with higher cognitive load, suggesting a fundamental adaptation in motor-visual coordination during multitasking.

4.6 Subjective Measures

There is a significant increase in mental demand ($\chi^2(3{,}N{=}16)$ = 27.27), physical demand ($\chi^2(3{,}N{=}16) = 26.78$, $p < .001$) and time pressure ("hurried or rushed") ratings ($\chi^2(3{,}N{=}16) = 20.78$, $p < .001$) with cognitive load. Participants reported feeling significantly less successful in task completion as cognitive load increased ($\chi^2(3{,}N{=}16) = 9.87$, $p < .05$). Affective states were similarly impacted, with participants reporting feeling significantly less in control ($\chi^2(3{,}N{=}16) = 8.38$, $p < .05$) and less confident ($\chi^2(3{,}N{=}16) = 8.87$, $p < .05$) in their abilities as cognitive load increased. Despite a higher workload, participants reported feeling more active ($\chi^2(3{,}N{=}16) = 8.76$, $p < .05$) during the 2-back condition compared to easier conditions. These subjective findings complement our behavioral measures by demonstrating that increased cognitive load not only impairs touchscreen interaction performance but also substantially alters drivers' perceived task load, confidence, and emotional state during in-vehicle touchscreen interactions.

5 Discussion

The goal of this study was to investigate the impact of cognitive load on in-vehicle touchscreen interactions and driving behavior. Our findings reveal that during multitasking involving driving and touchscreen interaction, these activities mutually affect each other, degrading their respective performance metrics. Pointing throughput decreased by 58.1% (Section 4.3.1), while lateral driving deviation also showed significant changes (Section 4.2.1). Changes in cognitive load potentially lead to further redistribution of visual-motor resources during this multitasking, exacerbating these effects. Although driving performance did not show significant variations due to our experimental design prioritizing driving safety (Section 4.2.2), pointing throughput significantly decreased by 20.2% as cognitive load increased (Section 4.3.2). Analyzing the focus of visual attention, we found that as cognitive load increased, drivers reduced their visual distraction time by 26.3%, from 1207 ms to 889 ms, when switching from driving to touchscreen interaction (Section 4.4), potentially reducing the number of target selections completed during each gaze transition.

In this section, we first discuss how driving and touchscreen interaction mutually influence each other, then further examine how increasing cognitive load affects various tasks during multitasking. We explore interesting behavioral pattern changes that cognitive load and multitasking induce in users, and propose several design guidelines for in-vehicle touchscreens.

5.1 Mutual Influence Between Touchscreen Interaction and Driving

Previous works [38] have investigated how touch interaction affects driving performance. Our research further validates their conclusions while demonstrating that this influence actually exists bidirectionally, regardless of the driver's cognitive load level.

For driving performance, after touchscreen interaction was introduced, although throttle control variability showed no significant change, participants' steering wheel control

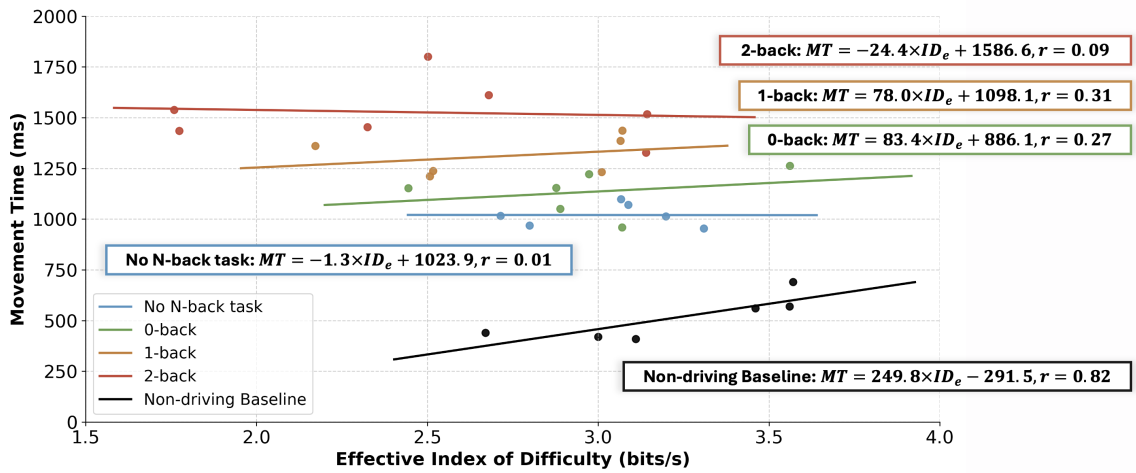

Figure 10: Participant 6's Fitts' law model fit. The black baseline represents tasks completed in a non-driving state.

significantly decreased, with lateral deviation from the road center increasing from 0.086 m to 0.122 m, revealing the potential dangers touchscreen interaction poses to driving safety.

In addition, we focused more on examining how driving degrades touchscreen pointing performance. Pointing throughput decreased dramatically by 58.1%, indicating that participants were not only mentally affected by driving but also physically performed clicking actions much slower.

To explain this phenomenon, we analyzed Fitts' law models for individual participants and surprisingly discovered that movement time no longer exhibited typical sensitivity to the effective index of difficulty (ID_e) across all cognitive loads when driving. As an example, Figure 10 shows P6's models. While movement time increased with rising cognitive load, it remained almost unchanged with rising ID_e. In contrast, P6's movement time increased with ID_e in the standard Fitts' law task without driving.

We propose that this phenomenon can be explained by participants' ballistic finger movements and visual attention allocation during pointing actions. As reported in Section 4.5, even without an *N*-back task, ballistic movements preceded gaze transitions to the touchscreen in 62.9% of pointing trials. This inverted motor-visual sequencing primarily explains the observed insensitivity of movement time to ID_e when driving. In contrast to standard Fitts' law paradigms that predominantly model motor control capabilities, our driving multitasking paradigm incorporated the process of visual attention-shifting and target acquisition into the recorded pointing movement time. These observations align with what we noticed during the experiment—participants' finger movements often exhibited very brief hovering and hesitation during ballistic movement, transforming the process from a "rapid ballistic movement phase" into a multi-step process waiting for visual feedback to correct the motor system. This "hand-before-eye" mechanism also explains the observation of transient hand-hovering or circling behaviors, likely representing movement initiation based on prior motor experience, followed by mid-trajectory pauses awaiting visual confirmation after target acquisition.

5.2 Impact of Cognitive Load on Touchscreen Interaction and Driving

Most of our experimental results examined how different cognitive load levels further affect touchscreen and driving performance. Participants' mental resources indeed decreased under high cognitive load, but they prioritized driving safety, maintaining the driving performance while touchscreen interaction performance and the attention allocated to touch interaction significantly decreased.

For driving itself, we emphasized in instructions that participants should prioritize driving tasks as if they were actually driving rather than playing an immersive game. Consequently, driving lateral deviation and throttle control variability showed no significant differences, suggesting that participants were maintaining good driving performance as instructed.

On the other hand, touchscreen pointing performance showed significant degradation. Our *post hoc* pairwise comparisons of pointing throughput revealed that both movement time per touch and pointing throughput significantly decreased under high cognitive load. As the cognitive load increased, the "hand-before-eye" pattern became more pronounced. Under 2-back conditions, the average single pointing duration even exceeded each on-screen glance duration, demonstrating participants' limited visual attention during target selection.

Higher cognitive load led to shorter visual distractions. Conversely, in the Fitts' study, pointing time actually increased as cognitive load increased. This opposite trend resulted in a reduction in the number of pointing trials participants could complete during each on-screen gaze period, from 1.30 times (SD = 0.41) without an *N*-back task to 1.09 times (SD = 0.66) with the 2-back task. These findings suggest that higher cognitive load limited drivers' ability to maintain prolonged visual attention and to perform consecutive pointing actions. Additionally, participants showed reduced engagement with the Fitts' task under increased cognitive load, suggesting a constraint on attentional resources.

As shown in Figure 9, high cognitive load reduced visual distraction duration, nearly eliminating prolonged distractions based on the 2-second safety rule. However, since driving safety remained consistent across load levels, fewer prolonged distractions do not necessarily indicate safer driving. This suggests the threshold for unsafe visual distraction may depend on cognitive load.

5.3 Design Guidelines For In-Vehicle Touch Interfaces

Our findings offer guidelines for designing cognitive load-aware adaptive touch interfaces. With the increasing availability of wearable sensors and in-cabin sensing technologies, real-time cognitive load estimation is becoming feasible. Below, we provide adaptive strategies based on specific empirical insights from our study, focusing on motor-visual behavior, physiological and behavioral cues, and interaction bottlenecks under multitasking pressure.

Detectable Signals of Cognitive Load via In-Vehicle Sensors. Beyond traditional physiological indicators such as pupil diameter and electrodermal activity (EDA), our study highlights a novel and actionable behavioral signal of cognitive load: the "hand-before-eye" coordination pattern. Under higher cognitive load, drivers frequently initiate ballistic hand movements before shifting their gaze. This anticipatory motor action becomes more prevalent as cognitive demands increase, suggesting it reflects an internal sense of time pressure. This behavior can be detected using lightweight sensing technologies, such as eye-tracking systems to monitor gaze direction and capacitive touch sensors on the steering wheel to identify hand-off events. By analyzing the relative timing between hand departure from the wheel and subsequent gaze transition to the screen, systems can infer real-time cognitive load. This approach offers a low-cost, scalable method for integrating behavioral markers into cognitive load estimation frameworks.

Adaptive Visual Search Efficiency. Our findings indicate that the primary bottleneck in touchscreen interaction under high cognitive load is likely not motor capability but rather the visual search latency required to locate targets after shifting gaze from the road to the screen. Increasing button size alone did not significantly improve performance, suggesting that physical targeting was not the dominant constraint. This result highlights the need for adaptive user interfaces that enhance visual saliency rather than relying solely on target enlargement. For instance, interfaces can dynamically increase perceptual prominence using contrast enhancements, ambient lighting adjustments, or subtle animations to draw attention to actionable elements. During cognitively demanding moments, systems could also re-prioritize interface layouts to elevate critical controls while suppressing less relevant options. Overly dense layouts should be avoided, as they may elevate search costs even if they optimize for screen space. Taken together, these findings call for a shift from size-focused designs to visually optimized interfaces that support fast, accurate target acquisition under multitasking.

Minimize Multi-Step Interactions During High Cognitive Load. Under high cognitive load, participants completed only 1.09 to 1.30 touchscreen interactions per glance, while movement times increased by 20.2% and visual distraction durations decreased by 26.4%. These findings indicate that drivers, when cognitively taxed, are unable to sustain multi-step interaction sequences within a single attention window. Interaction performance becomes constrained not by motor ability but by the temporal limits of safe visual disengagement from the road. To mitigate this limitation, interface designers should implement adaptive task flattening strategies that minimize the number of steps required to complete an interaction. For example, frequently used functions should be surfaced as one-tap actions when high cognitive load is detected. Additionally, user interface flows can be dynamically simplified (e.g., reducing nested menus or multi-stage confirmations) based on real-time estimates of cognitive demand.

Adaptive Alerts for Prolonged Gaze. The NHTSA's fixed 2-second off-road glance threshold is based on the assumption of uniform attentional capacity across all cognitive states. However, our findings show that under high cognitive load, drivers' off-road glances naturally shorten to an average of 889 ms, reflecting an implicit self-regulation strategy to maintain driving safety. This finding suggests that static thresholds may not adequately account for the dynamic nature of attentional resources. To address this, future driver-assistance systems should adopt cognitive load-sensitive thresholds for visual distraction warnings. When high cognitive load is detected, systems should initiate earlier warnings.

Non-Visual Feedback for Target Acquisition. The observation that participants' hands moved for targets before their eyes looked for those targets indicates they had prior expectations about the touchscreen and approximate target locations. For touch interaction tasks, in-vehicle touchscreens can incorporate haptic and audio feedback as indicators of target selection or provide cues for target acquisition, thereby reducing visual distraction. For example, auditory or tactile feedback can serve as click confirmation. Haptic feedback can also enable the hand to function as a sensor, effectively giving virtual buttons physical properties [82].

5.4 Generalizing N-back Tasks To Real Scenarios

Based on four different cognitive load levels in our study, we can generalize our experimental findings to real-world scenarios studied in previous research and quantify the cognitive load in these scenarios [25].

Specifically, driving without a N-back task can directly represent similar real driving scenarios. The 0-back task, which only involves hearing and repeating a number, parallels everyday activities such as listening to music or passive conversation. Our NASA-TLX mental load measurements and physiological signal analyses indicated no significant differences in cognitive load between having no N-back task and the 0-back task. Notably, even in these relatively low-demand conditions, on a 20-point scale, the mental load associated with continuous touchscreen interaction remains moderate ($M = 9.5$ without an N-back task, $M = 10.0$ with a 0-back task), highlighting the need for greater attention to the potential risks posed by excessive in-vehicle touch interaction.

The 1-back and 2-back tasks involve moderate and high-intensity memory functions, information processing, and retrieval, similar to real-world activities such as texting, phone conversations, or engaging in intense discussions. We observed a significant increase in cognitive load under these conditions, along with notable changes in pointing accuracy, throughput, and visual attention patterns. These findings suggest that touch interfaces and regulations for use under moderate to high cognitive load conditions may require specialized design considerations.

Beyond driving, we expect our results to extend to other high-demand, touch-based interaction contexts, such as factory operators monitoring equipment [54], people using phones while moving [59], surgeons interacting with touchscreen displays [17], or pilots and performance drivers [95] managing complex controls. In each context, divided attention similarly degrades touch accuracy and speed, highlighting the need for adaptive, context-aware interface designs.

6 Limitations and Future Work

Our study has several limitations that should be acknowledged. Participants were explicitly instructed to prioritize driving safety over touchscreen performance. While realistic, this instruction may have constrained our observations of trade-offs between driving and touchscreen tasks under varying cognitive loads. Moreover, even under low cognitive load, inattention remains dangerous—a concern that will grow as driving automation alters attention allocation.

Although our high-fidelity simulator afforded precise control over task parameters, it inevitably falls short of capturing the full complexity and emotional stakes of real-world driving. We simplified the scenario (minimal interactions, no other cars or pedestrians) to reduce extraneous variability, yet this choice limits ecological validity. Likewise, our ring-of-circle target arrangements, while convenient for controlled analysis, does not reflect the various layouts and presentations of production vehicle interfaces. Future research should embrace more naturalistic designs, for example by incorporating traffic lights, dynamic road scenarios, rich UI elements, and immersive settings (e.g., virtual reality, instrumented test tracks, a *Drive-In Lab* system [53]) to observe how drivers spontaneously allocate attention when mental demands are high.

Our multidimensional cognitive load assessment, which paired subjective (NASA-TLX) and physiological (pupil diameter and EDA) measures, provided complementary insights but was not without its challenges. EDA signals suffered noise from steering and hand movements, and pupil measurements are sensitive to ambient lighting. Integrating additional neural measures such as functional near-infrared spectroscopy (fNIRS) [29] or electroencephalography (EEG) [1] could provide more precise correlates of cognitive load.

Looking ahead, another promising avenue for future research is to examine individuals with extensive multitasking training, such as professional pilots or emergency response operators. These experts are known to operate effectively under high cognitive loads while performing complex tasks. Comparing their performance to that of general drivers could help isolate whether the performance degradation observed in touchscreen interactions and driving behavior is primarily driven by the inherent cognitive load or by limited multitasking ability.

Finally, our findings suggest potential benefits from cognitive load-aware interfaces, but we have not yet implemented or tested such adaptive systems. Future work should develop and evaluate interfaces that dynamically adjust based on detected cognitive load levels. These systems could incorporate real-time monitoring of physiological signals, hand and eye movements, or certain driving behaviors to dynamically modify interface elements, information density, presentation style, animation timing, and feedback modalities. User studies could assess how drivers respond to such adaptive interfaces over time and whether these interfaces enhance both safety and user experience.

7 Conclusion

This study quantifies how cognitive load affects drivers' touchscreen interactions and visual attention. It also quantifies how interacting with a touchscreen affects driving performance, and how driving affects touchscreen performance. We manipulated cognitive load using an auditory N-back task and verified successful manipulation through subjective responses, physiological responses, and interaction measurements. The concurrent performance of driving and touchscreen tasks demonstrates significant mutual interference, with driving causing a 58.1% reduction in pointing throughput and touchscreen use resulting in a 41.9% increase in lateral vehicle deviation. Higher cognitive loads significantly affected touchscreen interaction efficiency, with movement time increasing by 20.2% in the highest cognitive load condition and throughput decreasing by 16.9%, from 2.42 bits/s to 2.01 bits/s. Pointing accuracy was maintained at the expense of speed, resulting in this lower throughput. Drivers naturally adapted their visual strategies under increased cognitive load, reducing visual distraction durations by 26.4%, suggesting instinctive preservation of the primary driving task by limiting visual engagement with secondary tasks when cognitive resources were constrained. We observed more prolonged visual distractions under low cognitive loads, potentially creating safety risks despite drivers' perceived attentional capacity. A "hand-before-eye" movement pattern, where ballistic movements of the hand preceded gaze transitions, occurred in 71.9% of movements, offering an opportunity for low-cost detection of cognitive load levels.

This research advances our understanding of driver-touchscreen interactions and informs potentially safer in-vehicle interface designs by quantifying how cognitive load affects the interplay between cognitive demands, visual attention, and motor performance during driving and touchscreen pointing tasks.

Acknowledgments

This work was supported in part by the University of Washington Information School and by Toyota Research Institute (TRI). Any opinions, findings, conclusions, or recommendations expressed in our work are those of the authors and do not necessarily reflect those of any supporter.

References

[1] Pavlo Antonenko, Fred Paas, Roland Grabner, and Tamara Van Gog. 2010. Using electroencephalography to measure cognitive load. *Educational psychology review* 22 (2010), 425–438.

[2] Seyed Amir Hossein Aqajari, Emad Kasaeyan Naeini, Milad Asgari Mehrabadi, Sina Labbaf, Nikil Dutt, and Amir M Rahmani. 2021. pyeda: An open-source python toolkit for pre-processing and feature extraction of electrodermal activity. *Procedia Computer Science* 184 (2021), 99–106.

[3] Gayoung Ban and Woojin Park. 2024. Effects of in-vehicle touchscreen location on driver task performance, eye gaze behavior, and workload during conditionally automated driving: nondriving-related task and take-over. *Human Factors* 66, 12 (2024), 2651–2668.

[4] Stefan Becker, Parrish Hanna, and Verena Wagner. 2014. Human machine interface design in modern vehicles. *Encyclopedia of automotive engineering* (2014), 1–16.

[5] Kristen E Beede and Steven J Kass. 2006. Engrossed in conversation: The impact of cell phones on simulated driving performance. *Accident Analysis & Prevention* 38, 2 (2006), 415–421.

[6] Xiaojun Bi, Yang Li, and Shumin Zhai. 2013. FFitts law: modeling finger touch with fitts' law. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1363–1372.

[7] David P Broadbent, Giorgia D'Innocenzo, Toby J Ellmers, Justin Parsler, Andre J Szameitat, and Daniel T Bishop. 2023. Cognitive load, working memory capacity and driving performance: A preliminary fNIRS and eye tracking study. *Transportation research part F: traffic psychology and behaviour* 92 (2023), 121–132.

[8] Karel A Brookhuis, Cornelie JG van Driel, Tineke Hof, Bart van Arem, and Marika Hoedemaeker. 2009. Driving with a congestion assistant; mental workload and acceptance. *Applied ergonomics* 40, 6 (2009), 1019–1025.

[9] Duncan P Brumby, Dario D Salvucci, and Andrew Howes. 2009. Focus on driving: How cognitive constraints shape the adaptation of strategy when dialing while driving. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1629–1638.

[10] Mikołaj Buchwald, Szymo Kupiński, Adam Bykowski, Joanna Marcinkowska, Dawid Ratajczyk, and Marcin Jukiewicz. 2019. Electrodermal activity as a measure of cognitive load: A methodological approach. In *2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. IEEE, 175–179.

[11] Totsapon Butmee, Terry C Lansdown, and Guy H Walker. 2019. Mental workload and performance measurements in driving task: A review literature. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018) Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics 20*. Springer, 286–294.

[12] Jeff K Caird, Sarah M Simmons, Katelyn Wiley, Kate A Johnston, and William J Horrey. 2018. Does talking on a cell phone, with a passenger, or dialing affect driving performance? An updated systematic review and meta-analysis of experimental studies. *Human factors* 60, 1 (2018), 101–133.

[13] Council of European Union. 2019. Council regulation (EU) no 2019/2144. https://eur-lex.europa.eu/eli/reg/2019/2144/oj.

[14] ERFW Crossman. 1957. The speed and accuracy of simple hand movements. *The nature and acquisition of industrial skills* (1957).

[15] Edward Robert FW Crossman and PJ Goodeve. 1983. Feedback control of hand-movement and Fitts' Law. *The Quarterly Journal of Experimental Psychology* 35, 2 (1983), 251–278.

[16] Gianluca Di Flumeri, Gianluca Borghini, Pietro Aricò, Nicolina Sciaraffa, Paola Lanzi, Simone Pozzi, Valeria Vignali, Claudio Lantieri, Arianna Bichicchi, Andrea Simone, et al. 2018. EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Frontiers in human neuroscience* 12 (2018), 509.

[17] Roger D Dias, Minhtran C Ngo-Howard, Marko T Boskovski, Marco A Zenati, and Steven J Yule. 2018. Systematic review of measurement tools to assess surgeons' intraoperative cognitive workload. *Journal of British Surgery* 105, 5 (2018), 491–501.

[18] Nguyen Do Duc Khanh. 2022. Are touch screens the new mobile phone? The effects of touch screen devices in vehicles on distraction and driving behaviour. (2022).

[19] Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. 2017. CARLA: An open urban driving simulator. In *Conference on robot learning*. PMLR, 1–16.

[20] Sarah A Douglas, Arthur E Kirkpatrick, and I Scott MacKenzie. 1999. Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 215–222.

[21] Cognitive Workload While Driving. 2022. Multitasking While Driving: Central Bottleneck or Problem State Interference? (2022).

[22] David B Duncan. 1955. Multiple range and multiple F tests. *biometrics* 11, 1 (1955), 1–42.

[23] Johan Engström, Emma Johansson, and Joakim Östlund. 2005. Effects of visual and cognitive load in real and simulated motorway driving. *Transportation research part F: traffic psychology and behaviour* 8, 2 (2005), 97–120.

[24] Johan Engström, Gustav Markkula, Trent Victor, and Natasha Merat. 2017. Effects of cognitive load on driving performance: The cognitive control hypothesis. *Human factors* 59, 5 (2017), 734–764.

[25] Johan Engström, Christopher A Monk, Richard J Hanowski, William J Horrey, John D Lee, Daniel V McGehee, Michael Regan, Alan Stevens, Eric Traube, Marko Tuukkanen, et al. 2013. A conceptual framework and taxonomy for understanding and categorizing driver inattention. (2013).

[26] Abigail Evans and Jacob O Wobbrock. 2011. Input Observer: Measuring text entry and pointing performance from naturalistic everyday computer use. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. 1879–1884.

[27] Julian J Faraway. 2016. *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman and Hall/CRC.

[28] Fred Feng, Yili Liu, and Yifan Chen. 2018. Effects of quantity and size of buttons of in-vehicle touch screen on drivers' eye glance behavior. *International Journal of Human–Computer Interaction* 34, 12 (2018), 1105–1118.

[29] Frank A Fishburn, Megan E Norr, Andrei V Medvedev, and Chandan J Vaidya. 2014. Sensitivity of fNIRS to cognitive state and load. *Frontiers in human neuroscience* 8 (2014), 76.

[30] Paul M Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology* 47, 6 (1954), 381.

[31] Lex Fridman, Bryan Reimer, Bruce Mehler, and William T Freeman. 2018. Cognitive load estimation in the wild. In *Proceedings of the 2018 chi conference on human factors in computing systems*. 1–9.

[32] Milton Friedman. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association* 32, 200 (1937), 675–701.

[33] I Garcia, Sebastian Bronte, Luis Miguel Bergasa, Javier Almazán, and J Yebes. 2012. Vision-based drowsiness detector for real driving conditions. In *2012 IEEE Intelligent Vehicles Symposium*. IEEE, 618–623.

[34] Sergio Garrido-Jurado, Rafael Muñoz-Salinas, Francisco José Madrid-Cuevas, and Manuel Jesús Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (2014), 2280–2292.

[35] Rafael C Gonçalves, Courtney M Goodridge, Jonny Kuo, Mike G Lenné, and Natasha Merat. 2024. Using driver monitoring to estimate readiness in automation: a conceptual model based on simulator experimental data. *Cognition, Technology & Work* (2024), 1–16.

[36] Courtney M Goodridge, Rafael C Goncalves, Ali Arabian, Anthony Horrobin, Albert Solernou, Yee Thung Lee, Yee Mun Lee, Ruth Madigan, and Natasha Merat. 2024. Gaze entropy metrics for mental workload estimation are heterogenous during hands-off level 2 automation. *Accident Analysis & Prevention* 202 (2024), 107560.

[37] Courtney Michael Goodridge, Rafael Cirino Gonçalves, and İbrahim Öztürk. 2023. What do we mean by cognitive load? Towards more accurate definition of the term for better identification by driver monitoring systems. In *Adjunct Proceedings of the 15th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 256–259.

[38] Toru Hagiwara, Ryo Sakakima, Toshiro Hashimoto, and Toshitake Kawai. 2013. Effect of distraction on driving performance using touch screen while driving on test track. In *2013 IEEE intelligent vehicles symposium (IV)*. IEEE, 1149–1154.

[39] Joanne L Harbluk, Y Ian Noy, Patricia L Trbovich, and Moshe Eizenman. 2007. An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis & Prevention* 39, 2 (2007), 372–379.

[40] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[41] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[42] Dengbo He, Ziquan Wang, Elias B Khalil, Birsen Donmez, Guangkai Qiao, and Shekhar Kumar. 2022. Classification of driver cognitive load: Exploring the benefits of fusing eye-tracking and physiological measures. *Transportation research record* 2676, 10 (2022), 670–681.

[43] William S Helton and Katharina Näswall. 2015. Short stress state questionnaire. *European Journal of Psychological Assessment* (2015).

[44] Eckhard H Hess and James M Polt. 1964. Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 3611 (1964), 1190–1192.

[45] Sture Holm. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* (1979), 65–70.

[46] Hongwei Huang, Jiateng Li, Xuejing Feng, Jun Ma, and Bruce Mehler. 2025. Characteristics of Driver Peripheral Vision: How Drivers Respond to Ubiquitous Information on Wide-Area In-Vehicle Displays. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[47] International Organization for Standardization. 2002. Ergonomic Requirements for Office Work with Visual Display Terminals (VDTs) — Part 9: Requirements for Non-Keyboard Input Devices. Published February 15, 2002.

[48] Christian P Janssen, Duncan P Brumby, and Rae Garnett. 2012. Natural break points: The influence of priorities and cognitive and motor cues on dual-task interleaving. *Journal of Cognitive Engineering and Decision Making* 6, 1 (2012), 5–29.

[49] Anirudha Joshi, Aaditya Ajikumar, Kuldeep Umaraiya, and Prashik Chavan. 2022. Fitts' throughput and "absolute" finger precision by handedness, hand, digit and target width. In *Nordic human-computer interaction conference*. 1–11.

[50] Jahanvash Karim, Robert Weisz, and Shafiq Ur Rehman. 2011. International positive and negative affect schedule short-form (I-PANAS-SF): Testing for factorial invariance across cultures. *Procedia-Social and Behavioral Sciences* 15 (2011), 2016–2022.

[51] Tomas Knapen, Jan Willem de Gee, Jan Brascamp, Stijn Nuiten, Sylco Hoppenbrouwers, and Jan Theeuwes. 2016. Cognitive and ocular factors jointly determine pupil responses under equiluminance. *PloS one* 11, 5 (2016), e0155574.

[52] Mariska E Kret and Elio E Sjak-Shie. 2019. Preprocessing pupil size data: Guidelines and code. *Behavior research methods* 51 (2019), 1336–1342.

[53] Tuomo Kujala and Abhishek Sarkar. 2025. Evaluating In-Car Tasks' Distraction Effects with Drive-In Lab. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–24.

[54] Marta Lagomarsino, Marta Lorenzini, Elena De Momi, and Arash Ajoudani. 2022. An online framework for cognitive load assessment in industrial tasks. *Robotics and Computer-Integrated Manufacturing* 78 (2022), 102380.

[55] Dave Lamble, Matti Laakso, and Heikki Summala. 1999. Detection thresholds in car following situations and peripheral vision: Implications for positioning of visually demanding in-car displays. *Ergonomics* 42, 6 (1999), 807–815.

[56] Bidhan Lamichhane, Andrew Westbrook, Michael W Cole, and Todd S Braver. 2020. Exploring brain-behavior relationships in the N-back task. *NeuroImage* 212 (2020), 116683.

[57] John D Lee. 2007. Technology and teen drivers. *Journal of safety research* 38, 2 (2007), 203–213.

[58] Ja Young Lee and Willhelm E Lee. 2017. Modeling the effect of subtask boundaries on driver glance behavior. In *Driving Assessment Conference*, Vol. 9. University of Iowa.

[59] Tayla B Lino, Gabriella S Scarmagnan, Sidney A Sobrinho-Junior, Giovanna MF Tessari, Glaucia H Gonçalves, Hugo M Pereira, and Gustavo Christofoletti. 2023. Impact of using smartphone while walking or standing: a study focused on age and cognition. *Brain sciences* 13, 7 (2023), 987.

[60] Luca Longo, Christopher D Wickens, Gabriella Hancock, and Peter A Hancock. 2022. Human mental workload: A survey and a novel inclusive definition. *Frontiers in psychology* 13 (2022), 883321.

[61] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, et al. 2019. Mediapipe: A framework for building perception pipelines. *arXiv preprint arXiv:1906.08172* (2019).

[62] Jun Ma, Junyi Li, Zaiyan Gong, and Jihong Yu. 2017. *Impact of in-vehicle touchscreen size on visual demand and usability.* Technical Report. SAE Technical Paper.

[63] I Scott MacKenzie. 1989. A note on the information-theoretic basis for Fitts' law. *Journal of motor behavior* 21, 3 (1989), 323–330.

[64] I Scott MacKenzie. 1992. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction* 7, 1 (1992), 91–139.

[65] I Scott MacKenzie and Poika Isokoski. 2008. Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1633–1636.

[66] I Scott MacKenzie, Tatu Kauppinen, and Miika Silfverberg. 2001. Accuracy measures for evaluating computer pointing devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 9–16.

[67] I Scott MacKenzie and Aleks Oniszczak. 1998. A comparison of three selection techniques for touchpads. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 336–343.

[68] Stefan Mattes. 2003. The lane-change-task as a tool for driver distraction evaluation. *Quality of Work and Products in Enterprises of the Future* 57, 60 (2003), 7.

[69] Bruce Mehler, Bryan Reimer, and Joseph F Coughlin. 2012. Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Human factors* 54, 3 (2012), 396–412.

[70] Bruce Mehler, Bryan Reimer, Joseph F Coughlin, and Jeffery A Dusek. 2009. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation research record* 2138, 1 (2009), 6–12.

[71] Sean M Montgomery, Nitin Nair, Phoebe Chen, and Suzanne Dikker. 2024. Validating EmotiBit, an open-source multi-modal sensor for capturing research-grade physiological signals from anywhere on the body. *Measurement: Sensors* 32 (2024), 101075.

[72] Tina Morgenstern, Daniel Trommler, Frederik Naujoks, Ines Karl, Josef F Krems, and Andreas Keinath. 2023. Comparing the sensitivity of the box task combined with the detection response task to the lane change test. *Transportation research part F: traffic psychology and behaviour* 93 (2023), 159–171.

[73] Thierry Mousel and Alexander Treis. 2017. Hands Off Detection Requirements for UN R79 Regulated Lane Keeping Assist Systems. In *25th International Technical Conference on the Enhanced Safety of Vehicles (ESV) National Highway Traffic Safety Administration*.

[74] National Highway Traffic Safety Administration. 2012. *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices.* Tech Report. National Highway Traffic Safety Administration, Washington, DC.

[75] Donald A Norman and Stephen W Draper. 1986. *User centered system design; new perspectives on human-computer interaction.* L. Erlbaum Associates Inc.

[76] Hans-Christoph Nothdurft. 2000. Salience from feature contrast: Variations with texture density. *Vision research* 40, 23 (2000), 3181–3200.

[77] Joakim Östlund, Lena Nilsson, Jan Törnros, and Åsa Forsman. 2006. *Effects of cognitive and visual load in real and simulated driving.* Statens väg-och transportforskningsinstitut.

[78] Adrian M Owen, Kathryn M McMillan, Angela R Laird, and Ed Bullmore. 2005. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human brain mapping* 25, 1 (2005), 46–59.

[79] Julie Paxion, Edith Galy, and Catherine Berthelon. 2014. Mental workload and driving. *Frontiers in psychology* 5 (2014), 1344.

[80] Julie Paxion, Edith Galy, and Catherine Berthelon. 2015. Overload depending on driving experience and situation complexity: which strategies faced with a pedestrian crossing? *Applied ergonomics* 51 (2015), 343–349.

[81] Stephen J Payne, Geoffrey B Duggan, and Hansjörg Neth. 2007. Discretionary task interleaving: heuristics for time allocation in cognitive foraging. *Journal of Experimental Psychology: General* 136, 3 (2007), 370.

[82] Matthew J Pitts, Lee Skrypchuk, Tom Wellings, Alex Attridge, and Mark A Williams. 2012. Evaluating User Response to In-Car Haptic Feedback Touchscreens Using the Lane Change Test. *Advances in Human-Computer Interaction* 2012, 1 (2012), 598739.

[83] Gary B Reid and Thomas E Nygren. 1988. The subjective workload assessment technique: A scaling procedure for measuring mental workload. In *Advances in psychology*. Vol. 52. Elsevier, 185–218.

[84] Bryan Reimer. 2009. Impact of cognitive task complexity on drivers' visual tunneling. *Transportation Research Record* 2138, 1 (2009), 13–19.

[85] Douglas A Reynolds et al. 2009. Gaussian mixture models. *Encyclopedia of biometrics* 741, 659-663 (2009), 3.

[86] Kilseop Ryu and Rohae Myung. 2005. Evaluation of mental workload with a combined measure based on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics* 35, 11 (2005), 991–1009.

[87] I Scott MacKenzie. 2015. Fitts' throughput and the remarkable case of touch-based target selection. In *Human-Computer Interaction: Interaction Technologies: 17th International Conference, HCI International 2015, Los Angeles, CA, USA, August 2-7, 2015, Proceedings, Part II 17*. Springer, 238–249.

[88] Hans Selye. 1936. A syndrome produced by diverse nocuous agents. *Nature* 138, 3479 (1936), 32–32.

[89] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2009. Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on information technology in biomedicine* 14, 2 (2009), 410–417.

[90] Samuel Sanford Shapiro and Martin B Wilk. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52, 3-4 (1965), 591–611.

[91] R William Soukoreff and I Scott MacKenzie. 2004. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International journal of human-computer studies* 61, 6 (2004), 751–789.

[92] Anna Soveri, Jan Antfolk, Linda Karlsson, Benny Salo, and Matti Laine. 2017. Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic bulletin & review* 24 (2017), 1077–1096.

[93] Rainer Stiefelhagen, Michael Finke, Jie Yang, and Alex Waibel. 1999. From gaze to focus of attention. In *Visual Information and Information Systems: Third International Conference, VISUAL'99 Amsterdam, The Netherlands, June 2–4, 1999 Proceedings 3*. Springer, 765–772.

[94] David L Strayer, Jason M Watson, and Frank A Drews. 2011. Cognitive distraction while multitasking in the automobile. In *Psychology of learning and motivation*. Vol. 54. Elsevier, 29–58.

[95] Huimin Tang, Boon Giin Lee, Dave Towey, and Matthew Pike. 2024. The impact of various cockpit display interfaces on novice pilots' mental workload and situational awareness: a comparative study. *Sensors* 24, 9 (2024), 2835.

[96] Andrew J Tattersall and Penelope S Foord. 1996. An experimental evaluation of instantaneous self-assessment as a measure of workload. *Ergonomics* 39, 5 (1996), 740–748.

[97] Daniel Trommler, Tina Morgenstern, Elisabeth M Wögerbauer, Frederik Naujoks, Josef F Krems, and Andreas Keinath. 2021. The box task-a method for assessing in-vehicle system demand. *MethodsX* 8 (2021), 101261.

[98] Yi-Fang Tsai, Erik Viirre, Christopher Strychacz, Bradley Chase, and Tzyy-Ping Jung. 2007. Task performance and eye activity: predicting behavior relating to cognitive workload. *Aviation, space, and environmental medicine* 78, 5 (2007), B176–B185.

[99] Pamela S Tsang and Velma L Velazquez. 1996. Diagnosticity and multidimensional subjective workload ratings. *Ergonomics* 39, 3 (1996), 358–381.

[100] Pieter Vanneste, Annelies Raes, Jessica Morton, Klaas Bombeke, Bram B Van Acker, Charlotte Larmuseau, Fien Depaepe, and Wim Van den Noortgate. 2021. Towards measuring cognitive load through multimodal physiological data. *Cognition, Technology & Work* 23 (2021), 567–585.

[101] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17 (2020), 261–272. doi:10.1038/s41592-019-0686-2

[102] Björn NS Vlaskamp and Ignace Th C Hooge. 2006. Crowding degrades saccadic search performance. *Vision research* 46, 3 (2006), 417–425.

[103] Nikolai von Janczewski, Jennifer Wittmann, Arnd Engeln, Martin Baumann, and Lutz Krauß. 2021. A meta-analysis of the n-back task while driving and its effects on cognitive workload. *Transportation research part F: traffic psychology and behaviour* 76 (2021), 269–285.

[104] Scott Watter, Gina M Geffen, and Laurie B Geffen. 2001. The n-back as a dual-task: P300 morphology under divided attention. *Psychophysiology* 38, 6 (2001), 998–1003.

[105] Alan Traviss Welford. 1960. THE MEASUREMENT OF SENSORY-MOTOR PERFORMANCE: SURVEY AND REAPPRAISAL OF TWELVE YEARS'PROGRESS. *Ergonomics* 3, 3 (1960), 189–230.

[106] Brady T West, Kathleen B Welch, and Andrzej T Galecki. 2022. *Linear mixed models: a practical guide using statistical software.* Chapman and Hall/CRC.

[107] Christopher D Wickens and C Melody Carswell. 2021. Information processing. *Handbook of human factors and ergonomics* (2021), 114–158.

[108] Frank Wilcoxon. 1992. Individual comparisons by ranking methods. In *Breakthroughs in statistics: Methodology and distribution.* Springer, 196–202.

[109] Marc Wittmann, Miklós Kiss, Peter Gugg, Alexander Steffen, Martina Fink, Ernst Pöppel, and Hiroyuki Kamiya. 2006. Effects of display position of a visual in-vehicle task on simulated driving. *Applied ergonomics* 37, 2 (2006), 187–199.

[110] Jacob O Wobbrock, Kristen Shinohara, and Alex Jansen. 2011. The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 1639–1648.

[111] Sol Hee Yoon, Ji Hyoun Lim, and Yong Gu Ji. 2015. Perceived visual complexity and visual search performance of automotive instrument cluster: A quantitative measurement study. *International Journal of Human-Computer Interaction* 31, 12 (2015), 890–900.

[112] Shumin Zhai, Jing Kong, and Xiangshi Ren. 2004. Speed–accuracy tradeoff in Fitts' law tasks—on the equivalency of actual and nominal pointing precision. *International journal of human-computer studies* 61, 6 (2004), 823–856.

[113] Huiming Zhang, Yingshi Guo, Wei Yuan, and Kunchen Li. 2023. On the importance of working memory in the driving safety field: a systematic review. *Accident Analysis & Prevention* 187 (2023), 107071.

A APPENDIX: FEASIBILITY STUDY

We ran an initial feasibility study to test our ability to induce cognitive load and measure related changes in physiological arousal in a task that involved driving-like movements (i.e., wheel and pedal adjustments). We additionally did not constrain head movements or control for luminance, factors that can introduce artifacts into measures of physiological arousal (e.g., pupils, electrodermal activity). We ran a version of the box task [72, 97], a task introduced as a more robust and generalizable alternative to the lane-change task [68] for research on distracted driving. During this task we manipulated participants' cognitive load while recording motor movements (steering wheel and pedal positions) and levels of physiological arousal via pupillometry and measures of EDA.

A.1 Participants

Twenty-four participants (12 identifying as women, 12 identifying as men) were recruited from the local community. Participants were all above 18 years of age, had valid drivers' licenses (median 19 years of driving experience), and normal or corrected-to-normal vision. Data from three participants were omitted: two participants fell asleep during the task, and the eyetracker would not work for another. These participants' data were omitted before any data analysis. The study design was reviewed and approved by an institutional review board, and all participants provided informed consent.

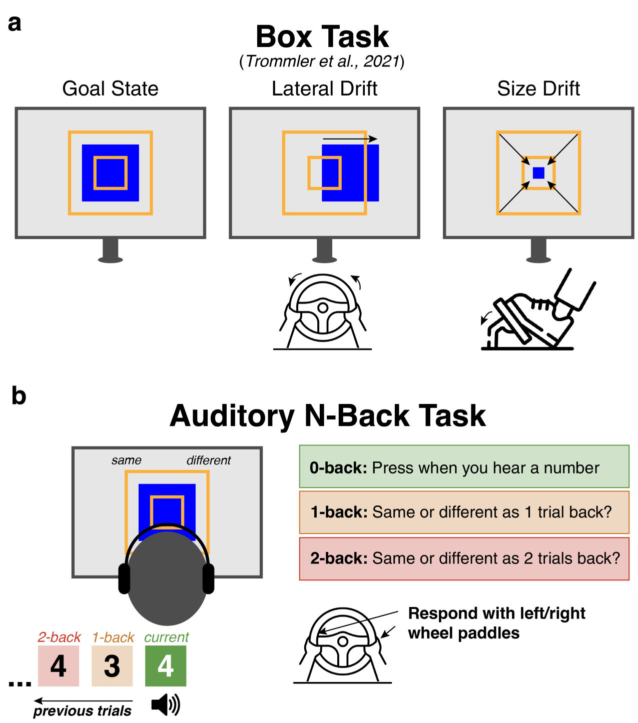

Figure 11: (a) Box task. (b) Auditory N-back task.

A.2 Materials and Methods

A.2.1 Box Task. We implemented a version of the Box Task [97] on a desktop computer connected to a 27" LCD monitor and a Logitech G29 Driving Force racing wheel and pedal setup. In this task, participants use the steering wheel and pedals to control the size and position of a square on a computer screen (Figure 11a). The goal of the Box Task is to position a blue square between a larger and smaller yellow square, such that the edges of the blue square do not move outside the boundary of the larger square or within the boundary of the smaller square. Throughout the task, the lateral position of the blue square slowly drifts left or right at a rate of 0.010 Hz. Concurrently, the size of the blue square drifts such that it grows or shrinks at a slightly faster rate of 0.125 Hz (for exact function controlling box position and size, refer to [97]). To keep the blue square in the goal position, participants turn the steering wheel to counteract the lateral drift and press either the accelerator or brake pedal to increase or decrease the size of square, respectively. Participants were given a series of instructions, practice trials, and a practice test to ensure that they understood the controls and the goal of the task before starting the main experiment.

A.2.2 Auditory N-back Task. Cognitive load was manipulated by having participants perform an auditory N-back task. In this task, participants heard an auditory sequence of numeric digits between 1 and 4 played in noise-canceling headphones and were asked to make a response using paddles at the back of the steering wheel based on numbers they heard n trials back. There were three conditions:

- **0–back:** press any paddle when a number is heard.
- **1–back:** press paddle to indicate whether the current number is the same or different from the number heard on the previous trial.
- **2–back:** press paddle to indicate whether the current number is the same or different from the number heard 2 trials back.

Because we only required same/different responses (rather than repeating the number head N-trials back), we restricted the number of digits to 4 to ensure a higher probability of "same" occurrences in the 1– and 2–back conditions (and avoiding high accuracy for simply answering "different" on every trial).

Participants completed blocks of N-back trials while concurrently performing the box task. Each block consisted of a sequence of 10 numbers with a 2.5-second inter-stimulus interval. Two seconds before the start of the block, text appeared above the large yellow square indicating the N-back condition that would follow. During the N-back trials, text appeared to the left and right of the large square indicating which paddle was to be pressed for "same" and "different" responses. The response order was randomized between N-back blocks.

Participants completed a total of 15 blocks of N-back trials (5 blocks of each N-back level, with the order randomized between participants). A 20-second break was given between each block of N-back trials.

A.2.3 Physiological measures. As participants completed the box task and N-back task, we recorded participants' gaze, pupil diameter, and EDA. Eye tracking was done with a desk-mounted EyeLink 1000+ (SR Research) with the Remote Camera upgrade to allow participants to move their head freely. Gaze and pupil diameter were recorded at 500 Hz. EDA was recorded using an Empatica E4 wristband sampling at 4 Hz.

A.2.4 Data Processing and Analysis. Pupil data were processed using a standard pupil processing pipeline [51]. Artifacts from blinks were removed from the pupil trace and missing samples were linearly interpolated. The pupil signal was then passed through a low-pass filter with a 10 Hz cutoff and the residual effects of blinks and saccades were removed using the `FIRDeconvolution` library [51]. Last, pupils were Z-scored within participants. For our subsequent analysis, pupil diameter during each N-back block was subtracted from a baseline period (2-second period before the start of the N-back block; Figure 14).

EDA was processed using the `pyEDA` package [2]. The raw EDA signal was decomposed into both tonic and phasic components. Similar to pupils, to assess changes in EDA induced by different levels of cognitive load, we subtracted baseline EDA recoded 2 seconds prior to each block from the EDA signal recorded during each N-back block.

A.3 Results

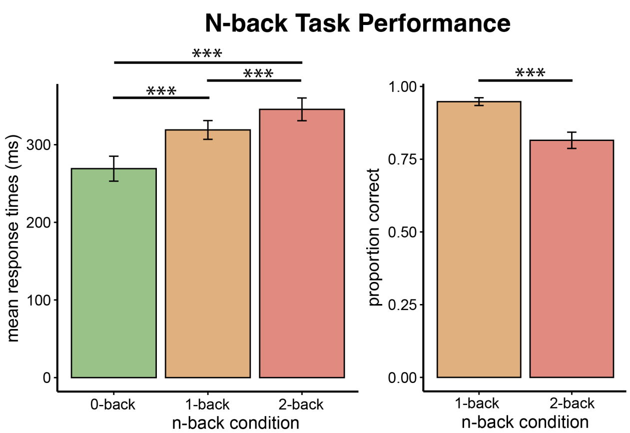

Figure 12: Participant response times and proportion correct scores on the N-back task. Accuracy is omitted for 0-back because there were no correct or incorrect responses. Bars indicate means and error bars indicate on standard error of the mean. ***:$p < .001$.

A.3.1 Performance decreases as N-back difficulty increases. To ensure that our cognitive load manipulation had the desired effect, we examined whether participant response times increased and accuracy decreased, with an increase in n on the N-back task. As expected, participants responded more slowly and less accurately as n increased (Wilcoxon signed-rank tests measuring differences in response times between N-back conditions: all $p < .004$; difference in accuracy between 1-back and 2-back: $p < .001$; Figure 12). These results indicate that participants experienced more cognitive load with greater n.

A.3.2 Cognitive load increases tonic physiological arousal. We next examined whether our cognitive load induction influenced physiological arousal responses measured by pupil diameter and EDA. Cognitive load had a robust influence on pupil diameter. As N-back increased, so too did average pupil diameter, with a particularly large difference in the 2-back condition compared to 0-back and 1-back (linear mixed effects model with random intercepts for each participant contrasting mean baseline-corrected pupil diameter between N-back conditions: 1-back vs 0-back: $\beta = 0.30$, $p = .005$; 2-back vs 0-back: $\beta = 0.93$, $p < .001$; 2–back vs 1–back: $\beta = 0.63$, $p < .001$; Figure 14a,b).

Cognitive load has a similar but less pronounced effect on slower, tonic changes in EDA (Figure 14c,d). Average changes in tonic EDA from baseline were higher in the 2-back compared to the 0-back task ($\beta = 0.35, p < .001$), but only marginally higher in the 2-back compared to the 1-back ($\beta = 0.18, p = .079$). There was no significant difference in tonic EDA between the 1–back and 0-back blocks ($\beta = 0.17, p = .107$). There were additionally no differences in average phasic EDA across N-back conditions (all $|\beta|$s< 0.14, all ps$> .206$).

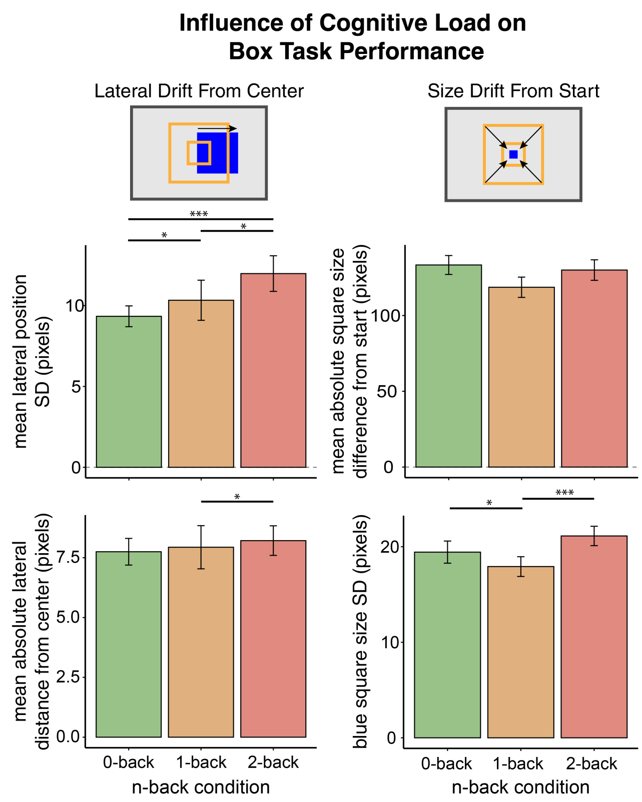

Figure 13: Influence of cognitive load on box task performance. Bars indicate means and error bars indicate on standard error of the mean. ***:$p < .001$, *: $p < .05$.

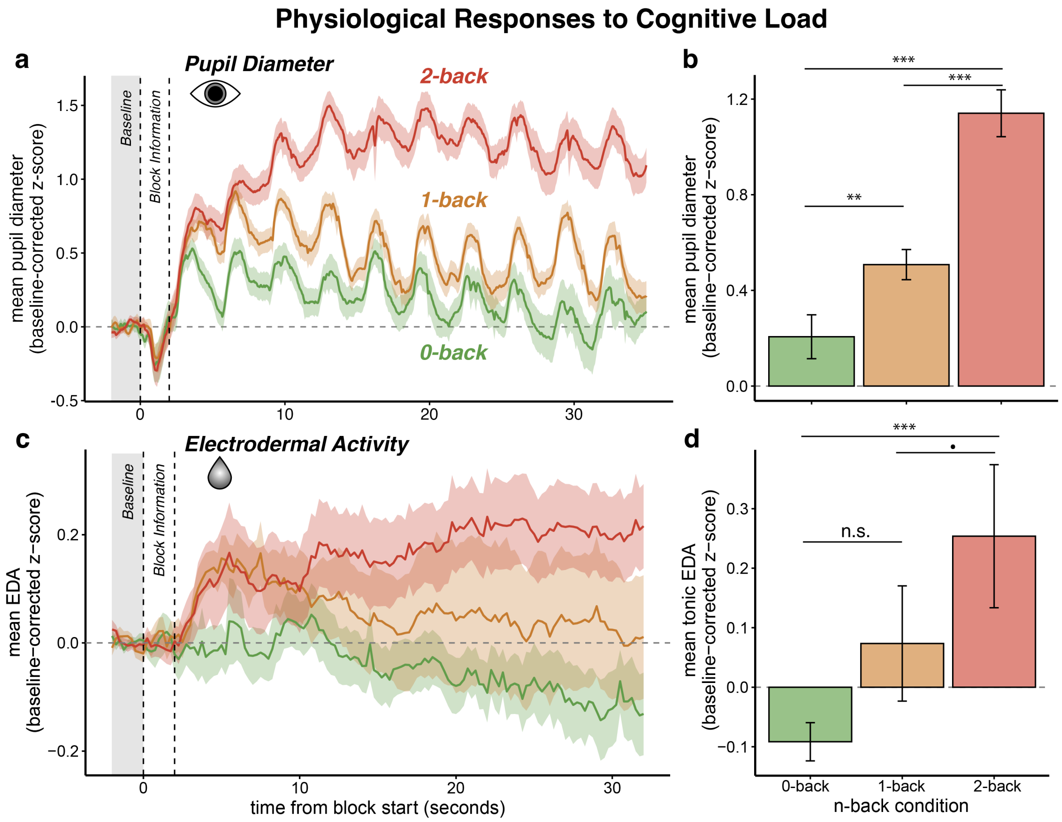

Figure 14: Influence of cognitive load as induced by an N-back task on physiological responses. (a, c) Changes in pupil diameter (a) and electrodermal activity (EDA; c) during the auditory N-back task. Solid lines in (a) and (c) correspond to between-subject means and shaded regions correspond to one standard error of the mean. (b, d) Average changes in pupil diameter (b) and tonic EDA (d) by N-back condition. Bars in (b) and (d) indicate overall averages and error bars indicate one standard error of the mean. ***: $p < .001$, **: $p < .01$, . : $p < .10$, n.s.: not significant.

A.3.3 Cognitive load has little impact on Box Task performance. We last explored whether increased cognitive load could be detected on the performance metrics of the box task. Overall, participant performance was quite high irrespective of cognitive load, with very few explicit positional errors (i.e., where the blue square touched either yellow square) committed across the experiment (median proportion time spent with blue square in error state [standard error of the mean] — 0-back: 0.014 [0.005], 1-back: 0.016 [0.010], 2-back: 0.025 [0.013]; signed rank tests comparing positional error rates across N-back conditions — all $ps > .118$).

Cognitive load had a small but inconsistent impact of people's ability to position the blue square in the goal position, even if explicit positional errors were low (Figure 13). As cognitive load increased, the absolute center position of the blue square tended to drift away from the center of the screen (linear mixed effects model contrasting the influence of different N-back conditions on log absolute lateral deviation from center — 1–back vs 0–back: $\beta = 0.09, p = 0.038$; 2–back vs 0–back: $\beta = 0.20, p < .001$; 2–back vs 1–back: $\beta = 0.11, p = .013$). However, cognitive load did not have a consistent influence on the variability of the blue square's position on the screen, measured by the average standard deviation of the square's position on the screen. Variability was slightly higher in the 2–back vs the 1–back condition (linear mixed effects model contrasting log standard deviation of square position across different N-back conditions — 2–back vs 1–back: $\beta = 0.85, p = .034$) but no different than the 0–back condition ($\beta = 0.50, p = .208$) or between the 1–back and 0–back conditions ($\beta = 0.34, p = .388$; Figure 13).

Cognitive load had less of an impact on people's ability to adjust the size of the blue square with the accelerator and brake pedals. We first measured people's tendency to let the size of the square drift away from its size at the start of the task, the optimal size to keep the square away from the edges of the yellow squares. Cognitive load did not influence participants' tendency to let the size of the square drift farther from the optimal size (linear mixed effects model contrasting log absolute deviation in square size compared to the start of the task — all $|\beta|s < 2.6$, all $ps > .365$; Figure 13). Variability in the size of the square did differ between N-back conditions, but not consistently with cognitive load. Size variability was higher in the 2–back condition compared the 1–back conditions, but not the 0–back condition and size variability was higher in the 0–back than the 1–back condition (linear mixed effects model contrasting log standard deviation of square size across N-back conditions: 2–back vs 1–back: $\beta = 0.17, p < .001$; 2–back vs 0–back: $\beta = 0.07, p = .117$; 0–back vs 1–back: $\beta = 0.09, p = .047$; Figure 13).

A.4 Discussion

The purpose of our feasibility study was to explore our ability to modulate cognitive load and explore its effects of different physiological responses and driving-related behaviors. Overall we successfully induced cognitive load with a N-back task as measured by increases in response times and decreases in accuracy in higher N-back conditions (Figure 12). We also found that cognitive load increased tonic arousal as measured by mean changes in pupil diameter and similar, albeit less pronounced changes in tonic EDA signals (Figure 14).

However, cognitive load had less of an influence on actual box task performance. Participants kept the blue square a little farther from the center and were slightly more variable with their steering wheel input as cognitive load increased. However, pedal interactions were less affected (Figure 13).

Previous studies using the box task and cognitive load manipulations add a tertiary Detection Response Task (DRT) in which a participant is required to provide a response (e.g., button press) to an unpredictable stimulus (e.g., haptic feedback) [72]. These studies find that cognitive load has a more robust influence on metrics of the DRT (e.g., response times, accuracy) than simple driving metrics. These results support the *cognitive control hypothesis* [24], which predicts that cognitive load primarily impairs motor responses that require higher cognitive control (e.g., responses to unpredictable stimuli) but not responses that are more automatic (e.g., standard steering wheel adjustments). This would suggest that efforts to detect and mitigate the effects of cognitive load on driving should likely focus on physiological measurements (e.g., pupils, EDA, eye-movements) and performance on tasks conducted while driving that require greater cognitive control (e.g., touch screen interactions).