Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predicting Driver Adaptability: A Hierarchical Computational Rationality Model of Bounded-Optimal Visual Sampling in Lateral and Longitudinal Vehicle Control

Toni Hautaoja

toni.n.a.hautaoja@jyu.fi

University of Jyväskylä

Tuomo Kujala

University of Jyväskylä

Ting Han Wei

Kochi University of Technology

Xiangshi Ren

Kochi University of Technology

Research Article

Keywords: attention allocation, computational modeling, driver behavior, occlusion, reinforcement learning, visual sampling

Posted Date: November 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10820650/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Attention allocation plays a central role in how drivers adapt to uncertainty resulting from human attentional and perceptual limitations, particularly when visual information must be sampled intermittently while maintaining control of the vehicle. However, how drivers adapt their visual sampling behavior when uncertainty arises simultaneously in several part-tasks of driving, such as lateral and longitudinal vehicle control, remains insufficiently understood. We developed a computational rationality model to predict drivers' adaptation of attention allocation in a driving task requiring lane keeping and speed maintenance under visual occlusion. Using reinforcement learning in a simulated environment, the model employs separate but interacting policies for lateral control, longitudinal control, and supervisory visual sampling of task-relevant information. Attention-dependent internal state estimates evolve as a function of intermittent sensory access, control actions and ongoing vehicle motion and can increasingly diverge from the external environment. In contrast to earlier work, the present model extends the lateral control task to include continuous speed regulation and evaluates the integrated model on a curved road with sinusoidal elevation changes. The model showed good aggregate correspondence with human data collected from 30 participants in a driving simulator, with all eight evaluated model means falling within one human standard deviation of the corresponding human means. The model provides a formal account of how adaptive visual sampling may emerge when lane keeping, speed regulation and limited visual information must be balanced under uncertainty. More broadly, the framework supports extension to additional driving subtasks and future investigation of inter-individual differences in attentional adaptation.

1 Introduction

Driving depends on the coordination of visual information, internal predictions and vehicle control. Although drivers use vision to monitor task-relevant states such as lane position, road curvature and speed, these states do not need to be observed continuously. During brief interruptions in visual information, control can continue on the basis of an internal estimate of the vehicle and roadway state (Godthelp, 1986). As time passes without new visual input, this estimate becomes increasingly dependent on prediction and therefore increasingly uncertain (Senders et al., 1967; Kujala & Lappi, 2021). Visual sampling can therefore be understood as a temporal process in which drivers update task-relevant information while control continues under partial observability.

Driver distraction research has shown that diverting visual attention from the roadway for too long can impair access to task-relevant information, and interactions with in-vehicle technologies can therefore have negative effects on driving performance and safety (Horrey et al., 2006; Regan & Oviedo-Trespalacios, 2022). However, continuous visual attention to every driving-relevant state is impossible because multiple task-relevant states must be monitored even within the driving task alone while the human visual processing capacity is limited. Fortunately, continuous visual attention is not required for safe driving if the required information is sampled sufficiently (Kujala et al., 2023; Godthelp, 1986). Redundant visual sampling of a state with minor or infrequent changes would also be a waste of

cognitive resources. Hence, the present question is not simply whether drivers look away, but how visual information is sampled between task-relevant states when control continues between observations.

The visual occlusion method, originally introduced by Senders et al. (1967), has provided a useful way to study this process by allowing drivers to determine how long they are willing to drive without visual information. In the self-paced version of the method used here, the driver's view is intermittently blocked using occlusion goggles or displays and the driver can end the occlusion by requesting visual information. Self-selected occlusion times can be interpreted as subjective estimates of spare visual capacity, that is, the time available for attention to be directed elsewhere while still maintaining sufficiently safe driving performance (Kujala et al., 2023). Supporting the broader relevance of this measure, drivers' occlusion time preferences have been shown to relate to their off-road glance durations in unstructured secondary in-vehicle tasks that were low in visual and cognitive demand, although this relationship becomes weaker when the secondary task imposes stronger demands of its own (Grahn et al., 2023).

More recently, occlusion has been reviewed as a method for assessing the attentional demand of driving tasks, thereby shifting the focus from merely describing observed gaze behavior toward estimating when visual information is required for safe performance (Kujala et al., 2023). This perspective is consistent with the concept of appropriate uncertainty: rather than requiring continuous attention, safe driving depends on driver keeping uncertainty about task-relevant states within acceptable limits (Kujala & Lappi, 2021). The underlying assumption is that instead of maintaining continuous, resource-wasting visual attention to every task-relevant state, drivers regulate their visual sampling. Driving performance and safety are then dependent on the timing and frequency of visual sampling to keep uncertainty about internal task-state estimates at an appropriate level.

Recent empirical and computational work has clarified this uncertainty management problem in lateral vehicle control. In an occluded lane keeping task, drivers' visual sampling behavior has been shown to vary with speed, lane position, time-to-line-crossing at the onset of occlusion, and steering during occlusion, although stable individual preferences also play a major role in self-selected occlusion times (Grahn et al., 2024). Complementing this empirical work, Hautaoja (2026) developed a computational rationality model in which a reinforcement learning agent learned when to sample visual information while keeping the sampling minimal and the vehicle within lane boundaries under partial observability. This showed that adaptive visual sampling in lane keeping can be modeled as bounded-optimal information acquisition under uncertainty and that inter-individual differences could be represented through psychologically interpretable parameters such as cautiousness. Complementing these empirical and computational approaches, Sarkar and Kujala (2025) developed a normative mathematical model for estimating situation- and driver-specific spare visual capacity in lateral vehicle control for safe driving.

However, lane keeping is only one component of the control problem faced by a human driver. Driving also requires longitudinal regulation, including the maintenance of appropriate speed or headway, and

intermittent visual sampling has been studied in this domain as well. Pekkanen et al. (2018) proposed a computational model of intermittent visual attention and longitudinal control in a car-following task, in which behavior was based on an internal state estimate updated through intermittent perceptual samples. Their model reproduced the relationship between time headway and visual attention and interpreted attention allocation as control of action uncertainty, which refers to the driver's uncertainty about whether the currently estimated traffic state is still accurate enough to guide safe longitudinal control without taking a new visual sample. Similarly, Kujala and Sarkar (2024) modeled spare visual capacity for safe driving in dynamic car-following situations and showed that the spare visual capacity in this task varies with situational factors such as distance headway and relative speed.

These studies establish longitudinal control as another important domain in which visual sampling can be understood as uncertainty management, but they treat it largely separately from lateral vehicle control. This separation is a simplifying assumption rather than a property of real driving. Lateral and longitudinal control are coupled through the vehicle dynamics and through drivers' regulation of safety margins (van Winsum & Godthelp, 1996). In curve driving, for example, drivers reduce speed when steering demands increase, compensating for larger steering errors so that time-to-line-crossing remains approximately stable across curve radii (van Winsum & Godthelp, 1996).

Thus, speed is not merely an additional task variable to be controlled alongside lane position, as it also changes the dynamics and safety margins of lateral control itself. Computational driver-control models have previously combined lateral and longitudinal control, for example by regulating steering together with throttle and braking commands in path-following tasks (Guo et al., 2004; Chatzikomis & Spentzas, 2009). However, these models have primarily addressed vehicle control performance rather than the driver's decision of when each task-relevant state should be sampled under uncertainty and limited attentional capacity.

This leaves an important gap in current driver attention modeling. Existing models have addressed visual sampling in lateral and longitudinal driving subtasks separately, and other models have combined lateral and longitudinal control without modeling the visual sampling decisions. Such models have not provided a formal account of how visual sampling should adapt when uncertainty develops simultaneously in lateral and longitudinal vehicle control and changes in one control process affect the other. Extending the problem from lane keeping alone to combined lane keeping and speed maintenance therefore introduces a broader attention allocation problem. Under intermittent visual access, the driver must no longer decide only when visual information is needed, but also which task-relevant state should be updated when multiple sources of uncertainty evolve in parallel. The attentional problem thus becomes supervisory; a limited visual resource must be allocated across interacting control processes whose demands may change over time.

We address this problem through the framework of computational rationality, in which human decision making and the resulting behavior are modeled as bounded-optimal adaptation to task constraints, perceptual limitations and task reward structures (Howes et al., 2009; Lewis et al., 2014; Gershman et al.,

2015; Oulasvirta et al., 2022). Computational rationality builds on bounded rationality (Simon, 1955) by making the agent's goals, information-processing limitations and environmental constraints explicit and asking what kind of behavior is optimal within those bounds.

Under visual occlusion, driving can be formalized as a partially observable decision problem, because the agent must act on incomplete information and rely on internal estimates of task states (Kaelbling et al., 1998). Reinforcement learning provides a method for optimizing such adaptive policies through interaction with a simulated environment (Sutton & Barto, 2018). Recent work has similarly described computational rationality as combining psychologically motivated assumptions with learned policies (Chandramouli et al., 2024). This approach builds on earlier computational rationality models of driving multitasking (Jokinen et al., 2021) and our previous model of visual sampling in lateral vehicle control (Hautaoja, 2026).

In the present study, we extend the previous lateral control computational rationality model into a hierarchical model of visual attention allocation in combined lateral and longitudinal vehicle control. The model includes separate but interacting policies for lateral control, longitudinal speed control and supervisory visual sampling of lane- and speed-related information. In contrast to the earlier lane keeping-only task (Hautaoja, 2026), the present environment requires continuous speed maintenance in addition to lane keeping, creating a control setting in which lateral and longitudinal task states may become uncertain in parallel under intermittent visual access. The model is designed to explain and predict how adaptive visual sampling may emerge when the driver must balance lane keeping, speed regulation and the allocation of limited visual information across such interacting task demands.

The hierarchical model architecture was designed to be extensible and generic in such a way that any other part-task of driving (e.g., headway maintenance, lane change, hazard perception) could be added to the model by defining the new task and adding it to the model as a sub-component. This extension contributes to driver attention modeling in three ways. First, it moves from a single-control-dimension account of visual sampling to an integrated lateral-longitudinal task. Second, it formalizes visual sampling as a supervisory decision over interacting control processes rather than as an isolated response to lateral uncertainty. Third, it provides a computational framework for simulating attentional adaptation in more ecologically valid driving tasks, with potential future applications in context-sensitive driver attention modeling, inter-individual differences in uncertainty management and the assessment of task-relevant internal representations in relation to driver awareness and engagement.

2 MATERIALS AND METHOD

Empirical data and preprocessing

For the empirical basis of the model, our research group conducted a driving simulator experiment at the University of Jyväskylä's medium-fidelity driving simulator located in the Faculty of Information Technology (see Fig. 1). The simulator has three 40" Samsung LED displays (4,320 × 900), a Logitech G25 steering wheel and pedals, an adjustable driver's seat, and a 2-DOF motion platform (CKAS T2s).

However, the motion was disabled in our experiment to prevent vestibular cues that are more demanding to model than visual sampling.

The experiment included 30 participants and followed a within-subject design. Participants had a mean age of 30.6 years (*SD* = 8.17, range = 19–47 years), a mean driving experience of 12.4 years (*SD* = 8.07, range = 2–30 years) and the sample included 15 women and 15 men. Participants drove in a curved and hilly road environment while wearing mobile eye-tracking glasses (Ergoneers Dikablis Glasses 3). The task required concurrently maintaining position within the center lane on a 3-lane road while keeping speed within the target range (100–110 km/h) under self-paced visual occlusion. During the task, participants were occluded by default and could request brief 0.5 s unoccluded glances by pulling a lever behind the steering wheel when they felt that visual information was needed for safe control. To assess their subjective estimates of spare visual capacity, their task was to try to maximize visual occlusion duration while staying on their lane and keeping their speed within the defined range.

Two complementary data streams were collected: driving simulator data and eye-tracking data. The simulator data was processed to obtain vehicle state and performance variables, including vehicle position, speed, steering, lateral offset, time-to-line-crossing (TLC), occlusion events, lane deviation indicators and speed deviation indicators. The eye-tracking data was processed using Ergoneers' D-Lab software (version 3.72). Areas of interest (AoIs) were manually defined for the road/lane area (center monitor) and the speedometer/speed information area (rightmost monitor). The relevant start and stop times of the driving clips were also marked manually so that the analyzed eye-tracking period corresponded to the actual driving task. The processed eye-tracking data were then synchronized with the driving simulator data using Syncster, a tailored software for synchronizing experimental data (version 1.0.0). This linked AoI glance data to simulator events such as occlusion onset and offset, driving time and vehicle-control measures. The synchronized data were analyzed in R (version 4.6.1) using RStudio version 2025.05.1 + 513.

The data were validated through sanity checks, including inspection of occlusion markers, timing alignment, missing values and consistency of the derived simulator and eye-tracking variables. The key driving-related variables included occlusion time (OT), TLC, speed maintenance, lane deviations, speed deviations, offset and steering amplitude. The key eye-tracking variables were the proportions of glance time allocated to three AoI categories: Lane, Speed and None. 'Lane' represented gaze directed at the road AoI, 'Speed' represented gaze directed at the speedometer AoI, and 'None' represented periods in which the display was occluded and neither visual channel was available. In the simulator, the speedometer was located sufficiently far away from the road view in order to get reliable difference between glances directed on the lane position and those directed on the speedometer during the unoccluded periods (see Fig. 2).

The resulting driving-performance variables and AoI ratios provided empirical reference measures and were used for descriptive assessment of human–model correspondence. This allowed us to evaluate whether the model reproduced both the control outcomes and the broad attention allocation pattern

observed in the human data. The goal was to give the model enough structure to learn the task reliably while avoiding an architecture or reward function that directly prescribed the target behavior. The model was therefore evaluated against human driving performance and AoI ratios, while the reward structure was kept relatively minimal so that the observed sampling behavior could emerge from task constraints, partial observability, prediction error and control failures.

Modeling approach and architecture

The present model extends the computational rationality-based model of visual sampling in lateral vehicle control developed by Hautaoja (2026), which itself built on the computational rationality approach to multitasking while driving introduced by Jokinen et al. (2021).

Whereas the previous model examined how an agent learns when to sample visual information in order to maintain safe lane keeping under partial observability, the present model extends the task to concurrent lateral and longitudinal vehicle control. The agent must regulate both its lane position and its speed while deciding whether visual information should be sampled from the lateral control channel, the longitudinal control channel or if neither is needed. The Python (version 3.9.6) source code was built on the earlier lateral control model and is available at https://gitlab.com/toninikolaialeksanteri/driver-adaptability.

The model was developed within the framework of computational rationality, which explains human behavior as bounded-optimal adaptation to the goals, constraints and information-processing limitations of the agent and its task environment (Howes et al., 2009; Lewis et al., 2014; Oulasvirta et al., 2022). The driving task was formalized as a partially observable sequential decision problem in which the agent must act under incomplete information and maximize long-term expected utility using attention-dependent internal estimates of the vehicle state (Kaelbling et al., 1998). When task-relevant visual information is unavailable, these estimates (i.e., beliefs, predictions) support continued control in the absence of direct observation. The implementation consists of five main components arranged in a hierarchical architecture (see Fig. 3): the external ground-truth driving environment, the internal attention-dependent belief model and three reinforcement learning policy agents:

- **a lane keeping agent**, which selected steering commands;
- **a speed maintenance agent**, which selected longitudinal speed control actions; and
- **a supervisory attention allocation agent**, which selected whether to sample lane-related information, speed-related information or neither.

Together, the two lower-level policies controlled the vehicle using features computed from the internal belief estimates (the policy inputs are features computed from those estimates), whereas the higher-level supervisory policy controlled their access to task-relevant visual information. This division reflects the distinction between continuous vehicle control actions and the allocation of limited perceptual resources. The architecture extends earlier hierarchical computational rationality models of multitasking in driving, in which supervisory control determines how attention should be allocated under uncertainty

(Jokinen et al., 2021) but differs in that the present model allocates attention between two interacting driving subtasks rather than between driving and an external secondary task.

The full model was implemented as an external ground truth environment coupled to an internal attention-dependent belief model. The external environment represents the true physical world, whereas the internal model represents the agent's current belief about the lane and speed states. The reinforcement learning agents receive these internal beliefs as observation inputs, while the model architecture prevents them from directly accessing the corresponding ground-truth states unless the relevant attention channel is actively sampled.

Simulated driving environment

The external driving environment represents the true physical world of the simulation. Among other things, it maintains the vehicle's actual position, heading, speed, steering state, lateral offset from the lane center, time-to-line-crossing and optional road-elevation variables. The simulation uses a fixed time step of 0.1 s, corresponding to a 10 Hz control loop. The main driving environment was a circular road with a radius of 1500 m and lane width of 3.75 m, matching the geometry used in the earlier empirical and computational work on occluded lateral vehicle control (Grahn et al., 2024; Hautaoja, 2026). Vehicle motion was represented using a simplified kinematic bicycle model with a wheelbase of 2.7 m. Such models retain the principal geometric relationship between steering and vehicle motion at substantially lower computational cost than models incorporating detailed tire dynamics (Kong et al., 2015). The vehicle was treated geometrically as a point for the purposes of lane-boundary crossings and time-to-line crossing calculations, as in the previous model. Steering commands were bounded to ± 0.1 radians, corresponding to approximately ± 5.7 degrees, which was based on the mean steering range observed in the human data.

The actions passed to the external environment consisted of a longitudinal speed change command and a steering command. The longitudinal command was implemented as a change in speed per simulation step rather than as acceleration directly expressed in m/s². This speed-control, or delta-v, command was clipped to ± 0.5 m/s per 0.1 s simulation step, equivalent to a maximum acceleration magnitude of approximately 5 m/s² over the time step. The external environment updated speed before updating position. The speed update can be summarized as:

$$\Delta v_{\mathrm{cmd}} = \mathrm{clip}\backslash\colon (\Delta v, -\Delta v_{\max}, \Delta v_{\max}),$$
$$v_{t+1} = \mathrm{clip}\ \left(v_t + \Delta v_{\mathrm{cmd}} + \Delta v_{\mathrm{noise}} + \Delta v_{\mathrm{grade}}, v_{min}, v_{max}\right),$$

where Δv is the raw longitudinal speed-change command, Δv_{cmd} is the command after clipping, $\Delta v_{\max} = 0.5$ m/s is the maximum permitted absolute speed change per simulation step, v_t and v_{t+1} are the vehicle speeds before and after the update, and $v_{\min}$ and $v_{\max}$ are the lower and upper speed bounds. The operator $\mathrm{clip}\,(x,\, a,\, b)$ restricts x to the interval $[a,\, b]$. $\Delta v_{\mathrm{noise}}$ denotes world-level speed noise and $\Delta v_{\mathrm{grade}}$ denotes the deterministic contribution of road grade. Logistic speed noise (0.05 m/s per dt) was added to the commanded delta-v at the external world-level. The actual applied

delta-v stored by the environment was the final speed change after delta-v command clipping (± 0.5 m/s per step), speed noise, hill effect and final speed-bound clipping.

Steering was also handled as a world-level executed control. The external environment started from the commanded steering angle, added logistic steering noise of 0.00075 rad per dt during training and 0.00025 during the model evaluation, and finally clipped the executed steering to ± 0.1 radians, which was based on the average steering variation observed in the human driving data. The noise values were selected through preliminary sensitivity testing, with higher training noise supporting robust learning and lower evaluation noise providing a better balance between control variability and empirical TLC correspondence. This noisy and clipped executed steering was stored as the applied steering value used in the external vehicle update. The position update used the same deterministic kinematic model in the true world update and in the internal forward simulation. If steering was zero, the vehicle moved in a straight line:

$$x_{t+1} = x_t + v_{t+1}\Delta\, t\backslash\text{:}\cos\left(\theta_{\,t}\right),\ y_{t+1} = y_t + v_{t+1}\Delta\, t\backslash\text{:}\sin\left(\theta_{\,t}\right).$$

Here, x_t and y_t denote the Cartesian vehicle coordinates at time step t, x_{t+1} and y_{t+1} the corresponding coordinates after the update, v_{t+1} is the updated vehicle speed, $\Delta\, t = 0.1$ s is the simulation time step and $\theta_{\,t}$ is the vehicle heading. For turning motion, R denotes turning radius, $L = 2.7$ m is the wheelbase, δ is the steering angle and $\Delta\,\theta$ is the change in heading over one simulation step. If steering was non-zero, vehicle motion was simulated along a circular arc with turning radius:

$$R = \frac{L}{\tan\left(\delta\,\right)}.$$

With this, heading changed according to:

$$\Delta\,\theta\ = \frac{v_{t+1}\Delta\, t}{R}.$$

The code first applies the delta-v command, speed noise, grade effect and speed limits, and then updates the vehicle position using the resulting speed. The external environment clamped the final position to world bounds, whereas the internal forward model did not apply this world boundary clamp. The environment also supported an undulating road profile that was made identical to the one used in the driving simulator experiment. This did not change the two-dimensional road geometry, but added a vertical sinusoidal elevation profile as a function of arc length along the road:

$$z\left(s\right) = A\backslash\text{:}\sin\left(\frac{2\pi\, s}{\lambda} + \phi\right),$$

where $z\left(s\right)$ is road elevation at arc-length position s, s is the distance along the road centerline, A is the elevation amplitude, λ is the wavelength and ϕ is the phase. In the present configuration, the

default hill amplitude was 6 m and the wavelength was 500 m. The implementation could adjust the effective wavelength so that an integer number of waves fitted into the closed loop, avoiding a discontinuity at the lap boundary. The local road grade was computed from the derivative of the elevation profile:

$$\frac{dz}{ds} = \frac{2\pi\, A}{\lambda} \cos\left(\frac{2\pi\, s}{\lambda} + \phi\right),$$

and the grade angle was:

$$\alpha\ = \arctan\left(\frac{dz}{ds}\right).$$

Here, the derivative dz/ds represents the local road slope and α is the corresponding grade angle. The grade affected speed through the deterministic term:

$$\Delta\, v_{\text{grade}} = -g \sin\left(\alpha\right) \Delta\, t,$$

where $g = 9.81$ m/s² is gravitational acceleration and $\Delta\, t$ is the simulation time step. Thus, uphill sections reduced speed and downhill sections increased it. This hill effect was included in both the external speed update and deterministic internal belief propagation when the road profile was assumed known. World-level speed noise was applied only in the external environment, not in the internal belief model. Hills were disabled during attention-policy training and enabled during final model evaluation, testing whether the learned supervisory policy transferred to longitudinal disturbances caused by the sinusoidal road profile.

TLC was used as the main lateral safety margin metric, following the definition introduced by Godthelp et al. (1984). In the model, TLC was estimated through deterministic forward simulation. The calculation assumed constant, unchanging speed and constant steering over the prediction horizon serving as a local risk proxy based on the current state. The forward simulation used an internal step size of $\frac{\Delta\, t}{10}$, with a maximum prediction horizon of 70 s, matching the simulator study. At each simulated sub-step, the predicted vehicle position was checked. TLC was returned as the simulated time until the vehicle first left the road. If no lane crossing occurred within the prediction horizon, TLC was returned as the maximum horizon.

Partial observability (internal beliefs)

Internal beliefs were represented as deterministic point estimates rather than probability distributions and therefore did not explicitly encode posterior variance, confidence or state uncertainty. The uncertainty and consequences of unavailable information were represented indirectly through belief drift and elapsed time since the last sample rather than through an explicit probability distribution over possible states. Separate belief states were maintained for lane- and speed-related information. The lane belief represented the model's internal estimate of the vehicle's road-relative state, including position,

heading, lateral offset and TLC, while the speed belief represented its estimate of vehicle speed. Both were maintained separately from the true vehicle state in the external environment.

To prevent ground-truth leakage, a hidden information channel could not access the corresponding external state variables. When a channel was visible, its belief state was synced with the external vehicle state. When it was hidden, the belief was propagated deterministically using the most recently commanded control actions and the same underlying transition dynamics as the external environment, with steering and speed noise disabled. The internally estimated state could therefore gradually diverge from the true state during occlusion, although this divergence (i.e., prediction error) was not explicitly represented as uncertainty. The available attention modes and corresponding commands for the attention allocation agent were:

$$\begin{aligned}
&\text{lane} \rightarrow \text{ lane\:channel\:visible,\:speed\:channel\:hidden,}\\
&\text{speed} \rightarrow \text{ speed\:channel\:visible,\:lane\:channel\:hidden,}\\
&\text{none} \rightarrow \text{ remain occluded, both\:channels\:hidden.}
\end{aligned}$$

When occlusion was disabled for training purposes, both channels were treated as visible. This setting was used during low-level lane keeping and speed maintenance training. During attention allocation training and simulation, occlusion was enabled, and the attention policy determined which channel was sampled. A visual sample, implemented as an unoccluded glance, lasted 0.5 s. Because the simulation time step was 0.1 s, each glance lasted five simulation steps. Glances were implemented as uninterruptible attention states. If the model was currently in a Lane or Speed glance, the glance counter advanced until the full glance duration had elapsed, after which attention automatically returned to the None state.

To enforce discrete, non-overlapping glance events, prevent the agent from exploiting unrealistically rapid switching between channels and avoid introducing an additional processing-delay parameter, the model could not switch directly from Lane to Speed or from Speed to Lane, and a new glance could begin only from the None state. The automatic end of a glance therefore produced at least one 0.1 s None state step before another glance could begin. It’s good to note, that human drivers were not constrained to maintain attention on a single AoI throughout the entire period. Hence, the model was a simplification in this regard.

A sampling initiation switch cost flag was activated when attention moved from None to either Lane or Speed. Therefore, the attention switching cost penalized active information sampling from occlusion, not passive continuation of a glance and not the automatic return from a glance to occlusion. When a new glance began, the newly visible channel was synchronized before the lower-level controllers acted on that simulation step. This allowed the newly visible channel to be updated immediately before the low-level controllers acted on that same time step. Without this synchronization, the low-level controller would act on stale beliefs, and the new sample would affect control only on the following step.

The internal belief update was synchronized with the external environment time step. Each external physics step triggered at most one belief update. At each step, the model checked which channels were currently visible and updated the internal beliefs according to the attention state. If both Lane and Speed were hidden, both beliefs were propagated deterministically using the previous speed belief, previous lane belief, last commanded steering and last commanded delta-v. When hills were enabled, the forward model also incorporated road grade effects based on the believed vehicle position and the assumed known road profile. Steering and speed noise were always excluded as unobserved world-level perturbations.

If Lane was hidden but Speed was visible, the speed belief was synchronized from truth, while the lane belief was propagated deterministically using the current speed belief, which had just been synchronized from the visible Speed channel and the last commanded steering. If Lane was visible but Speed was hidden, the lane belief was synchronized from truth, while the speed belief was propagated deterministically using the last commanded delta-v, and when hills were enabled, the road grade at the relevant lane-belief position. This allowed the internal model to account for estimated road slope without directly observing hidden speed. Believed TLC was calculated by combining the current lane-state belief with the current speed belief. Thus, when the Lane channel was visible, but the Speed channel was hidden, the lane-related state was synchronized with the external environment, while TLC was still calculated using believed speed. Consequently, the TLC prediction error recorded at a lane sampling event could reflect error in the predicted lane state, the predicted speed state or both.

When a channel became visible after being hidden, the model first recorded the pre-synchronization belief and compared it with the true sampled value. For lane sampling, this produced a sampling event comparing pre-sample belief TLC with true TLC. For speed sampling, it compared pre-sample speed belief with true speed. The resulting prediction errors were used for diagnostics and as event-based penalties in the attention allocation reward.

Policy observations

Here, beliefs refer to the model's internally maintained estimates of the task state, whereas policy observations refer to the task-specific inputs provided to each RL policy for action selection. These inputs were primarily derived from the internal beliefs and were supplemented by information about the time elapsed since the corresponding channel was last sampled.

2.1.1 Lane keeping observations

The lane keeping agent received a dictionary-form policy observation derived from the internal lane and speed beliefs and the elapsed time since the last lane sample. The observation included lateral offset as represented in the belief state, believed TLC clipped at 12 s and divided by 12, time since the last lane sample capped at 10 s and divided by 10, believed speed divided by the environment maximum speed, heading error divided by π and the normalized rate of change in believed lateral offset. The 12 s TLC cap was chosen based on the mean human TLC values observed in our previous study, which were approximately 12.4 s at 60 km/h and 10.4 s at 100 km/h (Hautaoja, 2026). It provided a bounded

normalization range while retaining greater resolution among lower TLC values associated with more immediate lane-departure risk. The cap was used as a pragmatic normalization bound rather than as a categorical safety threshold.

The rate of change in believed lateral offset was calculated as the difference between the current and previous believed offsets divided by the simulation time step. It was clipped to the range [− 2, 2] and divided by 2. For circular roads, heading error was calculated relative to the local circular tangent closest to the believed vehicle heading. For straight roads, it was calculated as the wrapped angular difference between the believed vehicle heading and the stored road heading. Although straight roads were not used in the present simulations, this calculation still remained part of the implemented observation logic of the model.

2.1.2 Speed maintenance observations

The speed maintenance agent received believed speed normalized by the environment's upper speed bound (120 km/h), signed believed speed error and time since the last speed sample. Speed error was expressed relative to the center of the target-speed band in units of the band half-width. The resulting value was clipped to [− 3, 3] and divided by 3, yielding an observation bounded to [− 1, 1]. Consequently, the lower and upper limits of the target band corresponded to − 1/3 and + 1/3 respectively, while values of − 1 and + 1 represented deviations of three band half-widths from the center. This scaling preserved graded information about speed deviations beyond the target band while limiting the influence of extreme values.

Although the speed-error input was centered on the midpoint of the target band, the final reward did not provide an explicit bonus for maintaining that exact speed. The input therefore did not explicitly impose a preference for the band center. However, a tendency to approach the center could still emerge if doing so provided a greater buffer against future deviations outside the rewarded speed range. Time-since-sample variables were also included in the lower-level observation dictionaries, but because occlusion was completely disabled during the separate lower-level training stages, these variables remained at zero during training and the lower-level policies were not explicitly trained to interpret nonzero times since the last sample.

2.1.3 Attention allocation observations

The attention allocation agent received belief-based state, risk and sampling-age inputs. These included TLC risk, signed speed error relative to the target-band center, time since the last lane sample and time since the last speed sample. The time-since-sampling inputs allowed the policy to distinguish recently sampled state estimates from beliefs that had been propagated for longer periods without direct observation. Both sampling-age variables were capped at 10 s and divided by 10. The TLC risk was calculated as

$$\mathrm{TLC}_{\mathrm{risk},t} = 1 - \frac{\mathrm{clip}\left(\mathrm{TLC}_{\mathrm{belief},t}, 0, 12\backslash\mathrm{hspace}\{0.17\mathrm{em}\}\mathrm{s}\right)}{12\backslash\mathrm{hspace}\{0.17\mathrm{em}\}\mathrm{s}},$$

where $\mathrm{TLC}_{\mathrm{risk},t}$ is the normalized TLC-risk input at simulation step t, $\mathrm{TLC}_{\mathrm{belief},t}$ is the currently believed TLC and 12 s is the normalization cap. A believed TLC of zero therefore produced maximum lane-related urgency, while TLC values at or above 12 s produced minimum immediate urgency. Signed speed error was calculated as

$$e_{v,t}^{\mathrm{AA}} = \mathrm{clip}\left(\frac{v_{\mathrm{belief},t} - v_{\mathrm{center}}}{\Delta\, v_{\mathrm{half}}}, -1{,}1\right),$$

where $e_{v,t}^{\mathrm{AA}}$ is the signed normalized speed-error input of the attention allocation agent, $v_{\mathrm{belief},t}$ is believed speed, v_{center} is the midpoint of the target-speed band and $\Delta\, v_{\mathrm{half}}$ is half the width of that band. The target-band limits therefore corresponded to − 1 and + 1 for the attention policy. All attention-policy inputs were derived from the internal belief state rather than from hidden ground-truth variables. See Appendix B for a summary table of all three model policies.

Reward structure

The model used separate reward functions for lane keeping, speed maintenance and attention allocation, allowing each lower-level policy to learn its control objective while the supervisory policy learned when task-relevant information should be sampled. The episode duration was 60 s. Successful completion of the episode (i.e., drive) produced a reward of 100 for each policy. Objective task failure occurred if the true vehicle state remained outside the lane boundaries or the permitted speed range for longer than the corresponding out-of-bounds timeout. These terminal conditions were always calculated from the external ground-truth state, even when the corresponding information channel was hidden from the policy. This is in line with a real-life situation where a driver can experience a crash, be penalized for speeding, or otherwise fail in speed maintenance even if not visually observing these events. The immediate reward received by each policy at simulation step t is denoted by $\mathrm{r_t}$.

2.1.4 Lane keeping reward

The lane keeping reward used believed TLC as the primary lateral safety margin signal, consistent with TLC as a time-related measure of lateral safety margin in driving (Godthelp et al., 1984; van Winsum & Godthelp, 1996). Following Hautaoja (2026), this was then scaled by believed speed relative to a fixed reference speed. The complete stepwise reward was thus defined as:

$$r_{\mathrm{lane},t} = \begin{cases} -100, & \tau_{\,\mathrm{lane,OOB},t} \geq \tau_{\,\mathrm{lane,max}}, \\ \mathrm{TLC}_{\mathrm{belief},t}\left(\frac{v_{\mathrm{belief},t}}{v_{\mathrm{ref}}}\right) + 100, & t \geq T_{\mathrm{episode}}\ \mathrm{without\backslash:terminal\backslash:failure}, \\ \mathrm{TLC}_{\mathrm{belief},t}\left(\frac{v_{\mathrm{belief},t}}{v_{\mathrm{ref}}}\right), & \mathrm{otherwise}. \end{cases}$$

Here, $r_{\mathrm{lane},t}$ denotes the lane keeping reward at simulation step t, $\mathrm{TLC}_{\mathrm{belief},t}$ is believed TLC, $v_{\mathrm{belief},t}$ is believed speed and $T_{\mathrm{episode}} = 60$ s is the maximum episode duration. The reference speed was $v_{\mathrm{ref}} = 60/3.6 = 16.67$ m/s. $\tau_{\,\mathrm{lane,OOB},t}$ denotes the continuous time spent outside the lane

boundaries and $\tau_{\mathrm{lane,max}}$ is the applicable lane out-of-bounds timeout. The timeout varied across the lane-training curriculum and was 3 s in the final training phase and model evaluation.

Unlike the TLC observation, which was capped at 12 s, the TLC value entering the lane reward was not capped and could reach the 70 s prediction horizon. This continuous stepwise reward encouraged the lane keeping policy to maintain a high believed lateral safety margin. If the true vehicle remained outside the lane for the applicable timeout, the episode terminated with a penalty of -100. Terminal failure also prevented the accumulation of subsequent stepwise rewards. If the episode reached the maximum duration without terminal failure, the completion reward of 100 was added to the lane reward on the final step. Thus, the continuous reward was calculated from the internal belief state, while objective lane failure was determined from the external ground-truth state and lane departure timer.

2.1.5 Speed maintenance reward

The reward treated speed regulation as a sparse, constraint-based objective: speeds within the instructed band were acceptable, while sustained deviations outside the band constituted task failure. Conceptually, this resembles safe RL formulations based on safety or performance requirements (García & Fernández, 2015), although the present implementation used ordinary terminal rewards and penalties rather than a formal constrained-RL algorithm. At ordinary nonterminal steps, the speed reward was zero. If true speed remained outside the target band continuously for at least 5.0 s, the episode terminated with a penalty. A completion reward was given if the maximum episode duration was reached:

$$r_{\mathrm{speed},t} = \begin{cases} -100, & \tau_{\mathrm{speed,OOB},t} \geq 5.0\,\mathrm{s}, \\ 100, & t \geq 60\,\mathrm{s\ without\ terminal\ failure}, \\ 0, & \mathrm{otherwise}. \end{cases}$$

Here, $r_{\mathrm{speed},t}$ denotes the speed-maintenance reward at simulation step t and $\tau_{\mathrm{speed,OOB},t}$ denotes the continuous time for which true speed has remained outside the target band. Although the terminal criterion was based on true speed, true speed was not exposed as a policy observation when the channel was hidden. During the attention allocation task, the frozen speed policy acted from its internal speed estimate, normalized speed-error input and time since the speed channel was last sampled. The sparse reward avoided directly prescribing a preferred speed within the acceptable range, such as the midpoint of the target band, allowing the target behavior to emerge from the specified bounds and task constraints.

2.1.6 Attention allocation reward

The attention allocation reward was also intentionally minimal to avoid prescribing the targeted human-like behavior to the model. Instead, it consisted of terminal lane failure and speed failure penalties, a completion reward, a sampling initiation switch cost for moving from None state to either Lane or Speed state, and event-based divergence penalties when hidden channels were sampled after occlusion. The switch cost represented the assumption that initiating a new visual sample is not free, conceptually consistent with evidence that switching between task sets carries measurable performance costs (Monsell, 2003; Kiesel et al., 2010), whereas the divergence terms represented sampling-revealed belief-

truth discrepancies or prediction errors between the hidden internal estimate and the state observed when visual information became available (Friston, 2010; Clark, 2013; Kujala & Lappi, 2021). These components of the attention allocation reward can be summarized schematically as:

$$\mathrm{r_{AA,t}} = \begin{cases} -300, & \mathrm{for\backslash:a\backslash:lane\backslash:OOB\backslash:termination,} \\ -100, & \mathrm{for\backslash:a\backslash:speed\backslash:OOB\backslash:termination,} \\ 100, & \mathrm{t} \geq\ 60\backslash\mathrm{hspace\{0.17em\}s\backslash:without\backslash:terminal\backslash:failure,} \\ -1.0 - 0.25\mathrm{D_{lane,t}}, & \mathrm{when\backslash:a\backslash:Lane\backslash:sample\backslash:is\backslash:initiated,} \\ -1.0 - 0.25\mathrm{D_{speed,t}}, & \mathrm{when\backslash:a\backslash:Speed\backslash:sample\backslash:is\backslash:initiated,} \\ 0, & \mathrm{otherwise.} \end{cases}$$

Here, $r_{\mathrm{AA},t}$ denotes the attention allocation reward at simulation step t, while $D_{\mathrm{lane},t}$ and $D_{\mathrm{speed},t}$ are the normalized Lane and Speed divergence terms defined below. The fixed value of 1.0 for the switch cost was applied only when the agent initiated a Lane or Speed sample from the None state, while the prediction error component with weight of 0.25 was applied only at the corresponding sampling event. Lane divergence was computed as the absolute difference between the believed TLC immediately before sampling and the true TLC observed at the sampling instant:

$d_{\mathrm{lane},t} = \left|\ \mathrm{TLC}_{\mathrm{belief},t^-} - \mathrm{TLC}_{\mathrm{true},t}\ \right|$. The raw divergence value was then normalized and capped as:

$$D_{\mathrm{lane},t} = \mathrm{clip}\left(\frac{d_{\mathrm{lane},t}}{12\backslash\mathrm{hspace\{0.17em\}s}}, 0, 1\right).$$

Here, $d_{\mathrm{lane},t}$ is the raw TLC belief–truth divergence at a Lane sampling event, $\mathrm{TLC}_{\mathrm{belief},t^-}$ is the believed TLC immediately before synchronization, $\mathrm{TLC}_{\mathrm{true},t}$ is the true TLC observed at sampling and t^- denotes the instant immediately before the sampled belief was synchronized with the true state. $D_{\mathrm{lane},t}$ is the corresponding normalized divergence, obtained by scaling the raw divergence by 12 s and clipping the result to $[0,\ 1]$. Speed divergence was computed analogously as the absolute difference between pre-sample believed speed and true speed at sampling:

$$d_{\mathrm{speed},t} = \left|\ v_{\mathrm{belief},t^-} - v_{\mathrm{true},t}\ \right|,$$

$$D_{\mathrm{speed},t} = \mathrm{clip}\left(\frac{d_{\mathrm{speed},t}}{3\backslash\mathrm{hspace\{0.17em\}m\,s^{-1}}}, 0, 1\right).$$

Here, $d_{\mathrm{speed},t}$ is the raw speed belief–truth divergence, v_{belief,t^-} is the believed speed immediately before synchronization and $v_{\mathrm{true},t}$ is the true speed observed at sampling. $D_{\mathrm{speed},t}$ is the corresponding normalized divergence, obtained by scaling the raw error by 3 m/s and clipping the result to $[0,\ 1]$. The default divergence weights were 0.25 for lane TLC divergence and 0.25 for speed divergence. These weights kept the divergence/ prediction error penalties smaller than terminal failure costs while still making large belief errors relevant for attention learning. The default switch cost was 1.0 per None-to-

Lane and None-to-Speed sample. Lane terminal failure was weighted more strongly than speed terminal failure in the attention reward, using a lane terminal multiplier of 3.0. This was a modeling choice intended to give sustained lane departure greater weight than sustained speed band deviation as lane departure can lead to more immediate and larger crash risk than speed band deviation. TLC served as the primary lateral safety margin measure (Godthelp et al., 1984).

With this, the attention reward did not directly encourage frequent sampling, too long occlusions nor a predefined Lane / Speed / None pattern. The attention allocation policy had to instead balance rewards for successful task completion against penalties for terminal control failures, sampling costs and prediction errors revealed when previously hidden channels were sampled. This reward design followed the broader modeling principle of the study, in which the empirical pattern of attention allocation served as an evaluation target while sampling behavior was allowed to emerge from task constraints, partial observability, prediction error and control demands.

Reinforcement learning implementation

The model policies were learned using reinforcement learning, which provides a method for approximating utility-maximizing behavior through repeated interaction with an environment (Sutton & Barto, 2018). RL was chosen for the present task because visual sampling and vehicle control involve sequential decisions whose consequences unfold over time under uncertainty. The aim was not to prescribe a fixed sampling rule or reproduce observed behavior through supervised prediction, but to examine whether adaptive control and attention allocation strategies could emerge from the specified task constraints, internal information limitations and reward structure. RL provided a suitable framework for identifying policies that balance immediate control demands, information acquisition costs and longer-term task success.

More specifically, all three policies were trained using the Stable Baselines3 (version 2.4.1) implementation of Proximal Policy Optimization (PPO) with MultiInputPolicy. PPO is an actor–critic policy-gradient method that uses a clipped surrogate objective intended to discourage excessively large destabilizing policy updates between successive iterations (Schulman et al., 2017; Raffin et al., 2021). We used the method to approximate control and sampling policies from the dictionary observations available to each agent. Partial observability was handled by the explicitly implemented internal forward model, which maintained attention-dependent point estimates of hidden lane and speed states.

The lane keeping and speed maintenance policies learned lower-level control strategies with full attention, while the attention allocation policy learned a higher-level sampling strategy over belief-based urgency and time-since-sampled inputs. Default PPO hyperparameters are reported in Appendix C. PPO provided model-free policy learning, meaning that the policies were optimized from experienced transitions and rewards without learning or using an explicit environment transition model for planning. However, the wider architecture also contained an explicit deterministic forward model for propagating hidden lane and speed beliefs. Hence, the implementation combined model-free reinforcement learning

at the policy level with an explicit deterministic forward model for internal state prediction during visual occlusion.

Training procedure

The model was trained hierarchically in three separate stages. Training began with simpler subproblems and gradually introduced more demanding conditions, supporting the stable acquisition of complex behavior (Bengio et al., 2009; Klink et al., 2022). This curriculum-based training approach allowed lower-level control competence to be acquired before training the supervisory attention allocation policy. The lane keeping policy was trained first, followed by the speed maintenance policy and finally the attention allocation policy with the two lower-level policies kept frozen. Detailed training procedures and curriculum configurations for all three stages are provided in Appendix A.

Model evaluation

The trained model was evaluated in simulation runs under fixed test conditions not used for policy updating. In the default simulation, the trained lane keeping, speed maintenance, and attention allocation policies were loaded. The pretrained lane-keeping and speed-maintenance policies acted deterministically. Attention actions were sampled stochastically from the learned policy distribution to generate policy-level variability across repeated simulation runs. The evaluation used a speed band of 100–110 km/h (the same as in the experiment), an initial speed of 100 km/h and the sinusoidal elevation profile. Each simulation episode lasted up to 60 s unless terminated earlier by lane or speed failure. During simulation, the model logged time, position, speed, heading, steering, TLC, offset, attention mode, and hill-profile info. The trace could be plotted and animated (see Fig. 4).

The code also recorded diagnostic belief-truth divergence measures. For lane sampling, it logged the belief TLC immediately before sampling and the true TLC at the sampling instant. For speed sampling, it similarly logged the believed speed immediately before sampling and the true speed at the sampling instant. Continuous hidden lane divergence traces were also logged and segmented by occlusion episode. These diagnostics could be exported as CSV files and plotted to examine how prediction error developed during occlusion.

The primary evaluation measures were mean TLC, mean OT, occlusion rate, mean speed, speed deviation rate, and the proportions of time allocated to Lane, Speed and None. For comparability with the empirical measure, TLC was operationalized as TLC at the subsequent Lane sampling event plus the duration of the preceding occlusion, providing an approximate TLC value for occlusion onset. Event counts were converted to events per 10 s whenever human and model driving durations were not identical:

$$\mathrm{Eventrate}_{10s} = \frac{N_{\mathrm{events}}}{T_{\mathrm{drive}}} \times \ 10\mathrm{s}.$$

The same operational definitions and analysis periods were used for the human and model data wherever possible. The final model configuration was simulated 30 times using different random seeds.

These runs represented stochastic realizations of the same trained policy rather than independently parameterized virtual drivers. Human and model distributions were summarized using the mean, standard deviation, median and range. For discrete or zero-inflated event measures, the interquartile range was also reported. Human-model correspondence was assessed descriptively using standardized distances expressed in units of the observed human standard deviation, following the general model-evaluation approach of Jokinen et al. (2021). For each outcome (j), the standardized distance between the model and human means was calculated as:

$$z_j = \frac{M_{\mathrm{model},j} - M_{\mathrm{human},j}}{SD_{\mathrm{human},j}}$$

Using the human standard deviation as the denominator expresses the distance between the model prediction and the center of the observed human distribution relative to the variability among human participants. Positive values indicate that the model mean was higher than the human mean, whereas negative values indicate that it was lower. For pragmatic descriptive interpretation in the present study, absolute standardized distances of $|z| \leq 1$ were classified as good correspondence, distances of $1 < |z| \leq 2$ as plausible but weaker correspondence and distances of $|z| > 2$ as clear deviations from the human data.

As an additional descriptive criterion, we examined whether each model mean fell within the observed minimum-maximum range of the human data. The mean absolute standardized distance across the outcomes was calculated as a compact descriptive summary. It was not treated as an inferential overall-fit statistic because the evaluated measures were not independent. Likewise, no independent-samples significance tests were conducted because repeated stochastic runs of a single trained policy are not equivalent to observations from independent human participants. Variation across simulation runs reflects stochastic within-policy variability rather than between-driver variability, which was not simulated.

3 Results

The final hierarchical model produced interacting visual sampling, lateral control and longitudinal control behavior across the 30 evaluation runs. Lane deviations were initially included as a performance measure but omitted from Table 1 because the final model produced none during evaluation, compared with 0.29 deviations per 10 s (*SD* = 0.22) in the human data. This greater model stability is considered further in the Limitations. Table 1 and Fig. 5 summarize the remaining human and model distributions together with the standardized model–human distances. The observation unit in the human data was the participant mean, while the model observation unit was the mean of one stochastic simulation run. Human driving-performance measures were available for all 30 participants. Valid eye-tracking and area-of-interest (AoI) data were available for 27 participants. All model measures were available for the 30 simulation runs.

Table 1
Human and model descriptive statistics and standardized model–human distances.

Measure	Human, *M* (*SD*)	Human, range	Model, *M* (*SD*)	Standardized distance, *z*
Time-to-line crossing (s)	8.92 (2.25)	5.94–13.97	9.57 (1.10)	0.29
Occlusion time (s)	0.85 (0.37)	0.23–1.83	0.97 (0.15)	0.31
Occlusions per 10 s	7.20 (2.23)	2.57–13.47	6.94 (0.68)	−0.12
Speed AoI ratio (%)	12.42 (6.27)	3.82–27.70	10.12 (1.54)	−0.37
Lane AoI ratio (%)	29.22 (6.26)	17.57–45.56	23.82 (3.15)	−0.86
None/occluded ratio (%)	58.36 (9.79)	34.21–77.29	65.90 (3.47)	0.77
Mean speed (km/h)	104.30 (2.52)	99.40–111.71	103.27 (0.44)	−0.41
Speed deviations per 10 s	0.53 (0.16)	0.14–0.81	0.41 (0.41)	−0.75

Note. Human (N = 30) for the driving-performance measures and (N = 27) for the AoI measures. Model (N = 30) for all measures. Standardized distances were calculated from unrounded values using the human standard deviation. Every model mean fell within the corresponding observed human minimum–maximum range.

Across the eight evaluated measures, all model means were within one human standard deviation of the corresponding human mean. None fell within the predefined range of plausible but weaker correspondence, and all eight model means fell within the observed human minimum–maximum ranges. The mean absolute standardized distance across the measures was 0.48 human standard deviations, indicating generally close descriptive correspondence between the final model configuration and the aggregate human data. In Fig. 5, positive values indicate higher model means and negative values indicate lower model means. The central shaded region between − 1 and + 1 represents good correspondence.

Visual sampling behavior

The model closely reproduced the temporal characteristics of human visual sampling behavior. The mean TLC was 9.57 s compared with the human mean of 8.92 s, resulting in $z = 0.29$. The OT measure showed a similar level of correspondence, with a model mean of 0.97 s compared with 0.85 s in the human data ($z = 0.31$). The model therefore remained occluded for slightly longer periods, although its mean was less than one third of a human *SD* from the human average and remained within the observed human range of 0.23–1.83 s. The model's occlusion frequency was also close to that observed in the

human data. It produced an average of 6.94 occlusions per 10 s compared with 7.20 occlusions per 10 s in the human data, corresponding to $z = -0.12$. The directions of the OT and occlusion frequency differences were internally coherent, with slightly longer occlusions accompanied by slightly fewer occlusions per 10 s. TLC, OT and occlusion frequency all remained within one human standard deviation of the corresponding human means, and overall the final model configuration reproduced the broad human pattern across these visual sampling measures (see Fig. 6).

The model also reproduced the broad division of visual attention between speed-related information, lane-related information and periods without visual information (see Fig. 7). Among the three attention allocation measures, the closest correspondence was observed for the Speed AoI ratio. Human participants allocated an average of 12.42% of the analyzed time to the speed-related AoI, while the model allocated 10.12%. This difference corresponded to $z = -0.37$. The model allocated slightly less time to the Lane category than the human participants. The human mean Lane ratio was 29.22%, whereas the model mean was 23.82%, resulting in $z = -0.86$.

This was the largest standardized model–human difference among the evaluated measures. The model mean nevertheless remained within one human standard deviation of the human mean and within the observed human range of 17.57–45.56%. Correspondingly, the model spent more time in the None (or occluded) state. The mean None ratio was 65.90% for the model and 58.36% for the human participants, producing a standardized distance of $z = 0.77$. Because the Lane, Speed and None ratios form a compositional measure that sums to 100%, the lower model Lane and Speed ratios and higher None ratio represent related aspects of the same attention allocation pattern. The main difference concerned the model's greater proportion of time under occlusion, although all three attention allocation measures remained within one human standard deviation of the corresponding human means.

Longitudinal driving performance

Longitudinal driving performance was also broadly consistent with the human data (see Fig. 8). The model's mean speed was 103.27 km/h compared with 104.30 km/h in the human data. This difference corresponded to $z = -0.41$, placing the model within one human standard deviation of the human mean. The model mean was also within the observed human speed range of 99.40–111.71 km/h. Variation in mean speed was smaller across the model runs than between the human participants, with standard deviations of 0.44 km/h and 2.52 km/h respectively. This difference was expected because the model runs represented repeated realizations of one trained policy and did not contain participant-specific parameter variation.

The model also produced a somewhat lower speed deviation rate than the human participants (see Fig. 9). Human participants produced an average of 0.53 speed deviations per 10 s, whereas the model produced 0.41 deviations per 10 s. The resulting standardized distance was $z = -0.75$, placing the model within one human standard deviation of the human mean. The model mean also remained within the observed human range. Variation in the model speed deviation rate was relatively large across runs, with

an *SD* of 0.41 compared with 0.16 in the human data, indicating that some simulation runs produced considerably more deviations despite the lower model average.

Prediction error development

To illustrate how prediction error developed while Lane information was unavailable, Fig. 10 shows believed and true TLC during a typical single lane-occlusion episode from an evaluation run. The two estimates remained relatively close during parts of the early occlusion but showed larger divergences as the episode progressed, with periods of partial convergence between larger prediction errors. Absolute TLC prediction error exceeded the 12 s cap used in the lane-divergence component of the attention reward at several points and remained above the cap immediately before Lane information was sampled again. The trajectory illustrates how deterministic belief propagation can produce substantial belief–truth divergence during occlusion, while also showing that the error does not necessarily increase monotonically with elapsed occlusion time, but depends on the evolving vehicle and predicted states. This diagnostic complements the aggregate human–model comparisons and was not included in the standardized distance calculations.

4 DISCUSSION

In this paper, we developed a hierarchical computational rationality model of visual sampling in combined lateral and longitudinal vehicle control. The model extended earlier computational rationality modeling of intermittent visual attention in lane keeping (Hautaoja, 2026; Jokinen et al., 2021) to a multitask setting in which the agent had to maintain lane position, regulate speed and decide when task-relevant visual information should be sampled. The model consisted of separate lower-level control policies for lane keeping and speed maintenance and a supervisory attention allocation policy that selected between lane sampling, speed sampling and continued occlusion. The central modeling assumption was that visual attention can be treated as adaptive information sampling under partial observability. Here, when visual information is unavailable, the driver or model acts from an internal belief state, and sampling becomes useful as internal estimates become older or task-related risk increases relative to the cost of acquiring and processing new information (Senders et al., 1967; Kujala & Lappi, 2021).

The results indicate that the hierarchical model configuration produced plausible aggregate behavior across visual sampling timing, supervisory attention allocation and longitudinal driving performance. All eight evaluated model means in the main fit table (Table 1) were within one human standard deviation of the corresponding human means and within the observed human ranges. Correspondence was particularly close for occlusion frequency, TLC, mean speed, occlusion time and the AoI ratio for speed. The largest remaining differences concerned the model's lower allocation of time to lane-related information, its greater proportion of time under occlusion and its lower average speed deviation rate. None of these differences exceeded one human standard deviation. However, this fit should be interpreted together with a qualitative limitation in lateral error production, as the current model

produced no lane deviations, whereas such events occurred steadily in the human data. Nevertheless, the results suggest that the model captured several key properties of human visual sampling and multitask control, while still overestimating the stability of lateral control.

Although the model did not explicitly represent state uncertainty as a probability distribution, its assumptions are broadly consistent with predictive coding and predictive processing accounts of cognition, in which behavior is guided by the interaction between internal predictions and sensory evidence as prediction error (Friston, 2010; Clark, 2013). In the present model, the internal lane and speed beliefs were propagated during occlusion and synchronized with the external environment only when the corresponding information channel was sampled (Pekkanen et al., 2018). The divergence between the internal belief estimates and the task-relevant environmental states observed at sampling can therefore be interpreted as a task-level prediction error. The model provides an apt mechanistic account of how temporary inattention to external visual information does not necessarily lead to a decrease in driving performance when the associated internal beliefs are sufficiently accurate, while becoming risky when prediction error grows too large and the model's internal estimates no longer support appropriate levels of control and certainty (Kujala & Lappi, 2021).

A key contribution of the model is that it treats visual sampling in driving as a supervisory multitasking problem following the supervisory multitasking approach of Jokinen et al. (2021). In many driving tasks, the driver does not only need to decide when to look, but also which information should be updated. The present model formalized this problem by allowing the supervisory policy to choose between lane information, speed information and to remain occluded. This extends earlier conceptual and experimental work on visual occlusion and spare visual capacity (see Kujala et al., 2023) by modeling attention allocation across multiple control-relevant information channels, shifting the focus from visual sampling as a single-channel timing problem toward visual sampling as a context-sensitive allocation problem between competing sources of uncertainty.

The results also support the broader computational rationality view that human-like behavior can emerge from task constraints, uncertainty, reward structure and limited perceptual access (Howes et al., 2009; Lewis et al., 2014; Oulasvirta et al., 2022). The model was not directly optimized to reproduce a predefined human attention pattern. Instead, the attention allocation policy had to balance successful task completion against terminal failures, sampling costs and the prediction errors revealed when hidden channels were sampled. This is important because it keeps the empirical data as an evaluation target rather than as a source of information for building the target behavior directly into the model. The plausible fit across several measures suggests that the model captured a useful approximation of how attention may be allocated when lateral and longitudinal control demands must be managed under intermittent vision.

At the same time, the model should be understood as a task-specific and mechanistic approximation rather than a complete theory of driver attention, serving as an initial foundational step. Validation against aggregate human data cannot prove that human drivers use the exact same internal

representations, reward functions or decision mechanisms (Wilson & Collins, 2019). The model instead demonstrates that a psychologically interpretable architecture based on internal beliefs, task constraints and costly sampling can reproduce several observed features of human behavior. In this sense, the model is both descriptive and explanatory, describing human-like visual sampling patterns at the aggregate level, while also explaining them as boundedly adaptive outcomes of the modeled task constraints.

The practical relevance of this approach is that the model provides a way to reason about attention and inattention in terms of context-dependent uncertainty instead of fixed glance-duration thresholds (Kujala & Lappi, 2021; Kujala et al., 2023). A two second glance away from the road and the lane keeping task may be relatively safe on a straight road at stable speed, but more problematic in a curve, on a hilly road or when speed is already drifting outside the instructed range. The present model captures this logic computationally by linking visual sampling to the internal belief state and the variable control demands of the task.

Limitations

Several limitations should be noted. First, as was already briefly mentioned, the current model produced no lane deviations in the evaluated runs, whereas human drivers showed a steady number of lane deviations, on average 0.29 (*SD* = 0.22) events per 10 s. This suggests that the model's lane control policy was too stable relative to human drivers. In practice, the learned steering policy and the lane-related reward structure may have suppressed rare lane departures, producing a floor effect in lateral-error outcomes. Thus, although the model reproduced several aggregate visual sampling and speed-related measures, it did not fully capture the variability and occasional control failures observed in human lane keeping. Future versions should therefore introduce more realistic sources of lateral-control variability, such as belief-related noise, perceptual uncertainty or participant-specific steering-skill or cautiousness parameters (Hautaoja, 2026), so that the rare but human-like lane deviation events can emerge without directly forcing them through the reward function. Also, the vehicle simulation dynamics were more realistic in the driving simulator than the simplified vehicle dynamics used in the model to reduce computational costs.

Although time-since-sampled variables were included in the lower-level (lane and speed agents') observation spaces, occlusion was disabled during standalone controller training. These inputs therefore remained at zero during pretraining and acquired non-zero values primarily when the frozen lower-level controllers were embedded in the integrated attention allocation task. This introduces a limited observation distribution shift between controller training and integrated evaluation. In future versions, the lower-level control policies could be trained with a broader distribution of inattention related states, so that they become more robust to the belief dynamics encountered during the full attention allocation task.

The attention dynamics also included a structural constraint, in which new Lane or Speed glances could begin only from the 'None' attention state. Once the new attention state was chosen, the model was

committed to that attention state for the full single sampling duration of 500 ms. This made the sampling process easier to control and interpret, but it represents a simplification of human gaze behavior. The model also treated Lane and Speed as mutually exclusive information channels, whereas human drivers may acquire some peripheral information from both the road and speedometer during an unoccluded glance. Human drivers can also shift gaze more flexibly between information sources, and real glances may overlap with transitional eye movements, peripheral preview and partial information pickup (Summala et al., 1996; Kujala et al., 2023). In the present model, the restriction helped maintain a clear distinction between active sampling and occluded belief propagation, but it also simplified the extent and temporal structure of human visual attention.

The current prediction error (i.e., divergence) penalty may not fully capture the development of prediction error during occlusion. As illustrated in Fig. 10, TLC belief–truth divergence can become substantial within a single occlusion, although its magnitude also varies non-monotonically as the vehicle and predicted states evolve. Because the TLC divergence term is capped, further increases in prediction error beyond the cap do not produce larger penalties. This may limit the extent to which the divergence penalty can shape longer-term Lane / Speed / None allocation patterns. Future versions should therefore explore divergence penalties that remain sensitive to larger prediction errors and better account for their temporal development during occlusion.

The model also relies on simplified belief-state representations and task-specific reward functions. The internal lane belief included variables such as TLC, offset and heading, while the speed belief represented the current speed state, from which its relation to the target speed band was derived. These representations are appropriate for the present task, but human drivers likely rely on richer and noisier perceptual-motor representations, including optic flow, peripheral vision, vestibular information, learned road expectations and uncertainty about their own control actions (Kujala & Lappi, 2021; Pekkanen et al., 2018; Salvucci & Gray, 2004). Similarly, the reward terms used in the model are psychologically interpretable, but their exact numerical values remain modeling choices. The switch cost, divergence weights, terminal multipliers and failure timeouts were chosen to produce stable learning and plausible behavior, not estimated directly from individual participants.

Another minor limitation concerns the preprocessing of the empirical data. The analysis used filtered data, and in some cases filtering removed entire occlusion events. This may have slightly affected the participants' measures, such as occlusion frequency, mean occlusion duration and the TLC values. The effect is expected to be minimal, but it should nevertheless be acknowledged when interpreting the exact numerical correspondence between the model and the human data. The model was also evaluated primarily at the aggregate level. Although the model means fell within the observed human ranges and within one human standard deviation for all eight standardized comparison measures, aggregate plausibility does not guarantee that the model captures the full distribution of individual attentional strategies. Some drivers may be more cautious, some more tolerant of uncertainty and some more sensitive to speed deviations or lateral risk. The present model therefore provides a plausible population-

level configuration, but a stronger test of the framework would require individualized parameter fitting and evaluation against participant-level patterns (Wilson & Collins, 2019; Hautaoja, 2026).

Finally, the model inherits the general limitations of reinforcement learning implementations of computational rationality. Training can be sensitive to reward balance, agents' inputs, hyperparameters and environmental details (Henderson et al., 2018). Although the present configuration produced plausible behavior across several variables, reinforcement learning does not automatically guarantee stable convergence or psychological interpretability. Computational rationality provides the theoretical principle that adaptive behavior emerges under bounds and task constraints, but RL is only one implementation method. Other approaches, such as inverse RL, hierarchical Bayesian modeling or cognitive architectures, may offer complementary and alternative ways to model attention allocation under uncertainty (Abbeel & Ng, 2004; Salvucci, 2006; Knill & Pouget, 2004; Körding & Wolpert, 2006).

Further research and implications

While addressing the aforementioned limitations, an important direction for future research is individualization. The present model was evaluated as a single hierarchical configuration against the human distribution. Future work could estimate participant-specific parameters related to cautiousness, control skill, belief noise and/or sampling cost. This would make it possible to create an ensemble of simulated drivers, each representing a different attentional strategy. Such a development would align closely with the broader idea of appropriate uncertainty, where safe and adaptive performance depends on the relation between individual safety margins, control abilities and environmental demands (Kujala & Lappi, 2021).

Another important direction is to extend the model to more complex and ecologically richer driving contexts. The present task included lane keeping, speed maintenance and intermittent visual sampling on a large-radius circular and undulating road, but real driving also involves traffic, headway regulation, overtaking, intersections, road users, weather, automation status and secondary tasks. Adding headway control would be a particularly natural extension, because longitudinal attention has often been studied in car-following contexts (Pekkanen et al., 2018; Kujala & Sarkar, 2024). More complex road and traffic environments would allow the model to estimate how attention allocation changes when several different sources of uncertainty compete for visual access. The current hierarchical architecture allows new subtasks to be added relatively easily, although interactions between such subagents would need to be explicitly encoded (cf. TLC belief vs. speed belief).

Third, future research could compare different computational implementations of the same theoretical idea. The current model uses reinforcement learning to implement computational rationality, but other methods could be used to infer or approximate rational information sampling behavior. Inverse reinforcement learning could help estimate implicit reward structures from human data (Abbeel & Ng, 2004). Bayesian models could provide a more explicit account of uncertainty representation in addition to the current point-estimate beliefs (Knill & Pouget, 2004; Körding & Wolpert, 2006). This could allow uncertainty itself to be examined as an internal model variable and provide a closer computational link to

predictive processing accounts of cognition (Friston, 2010; Clark, 2013; Kujala & Lappi, 2021). However, explicit uncertainty representation is not required for the present attention policy to develop adaptive sampling based on the consequences of continued occlusion. Additionally, cognitive architectures could provide more constrained assumptions about memory, timing and perceptual-motor processing (Salvucci, 2006). Comparing these different alternatives would clarify which aspects of the current results depend on the CR framework itself and which depend on the chosen RL implementation.

4.1.1 Implications for driver awareness and engagement

Beyond its role as a model of adaptive visual sampling, the present architecture also suggests a broader human–vehicle interaction perspective. By explicitly distinguishing the external task state from an attention-dependent internal belief state, the model provides a computational basis for asking not only where visual attention is directed, but also whether the driver maintains an adequate internal representation of task-relevant conditions. This perspective is compatible with the Human-Engaged Computing (HEC) framework, which emphasizes technological support for human awareness, agency and active engagement in addition to efficient task performance (Ren, 2016; Ren et al., 2019; Ren & Wang, 2026). In HEC terminology, this provides a potential computational starting point for examining aspects of Xin (□), understood here as subjective awareness and active cognitive engagement with the task. The present model does not directly measure such subjective states, but its belief-based architecture could support future work relating observable sampling behavior to the adequacy of the driver's internal representation of the driving situation.

From an applied perspective, the model might support development of more context-sensitive driver attention monitoring systems. Current systems often rely on simplified rules, such as warning the driver when gaze has been away from the road for too long. The present approach suggests that warnings should not depend only on elapsed glance time, but also on the uncertainty and control demands of the situation (Sarkar & Kujala, 2025). Looking away may be acceptable when the lane and speed states are stable, but unsafe when continued inattention is expected to compromise lane or speed control. In the long term, such models could support context-sensitive driver monitoring systems that estimate when inattention is becoming dangerous in a specific driving context. With sufficient empirical validation and participant-specific modeling, the same framework could also support cognitive calibration by estimating whether task-relevant information has been sampled sufficiently to maintain an adequate internal representation of the current driving situation, rather than relying on observable gaze behavior alone.

The model could also inform adaptive in-vehicle interface design. If a vehicle system can estimate when lane uncertainty or speed uncertainty is high, it could delay non-urgent notifications, simplify displayed information or avoid presenting visually demanding tasks during safety-critical moments. Such interfaces could also function as cognitive scaffolds, helping drivers update task-relevant information and maintain awareness and active engagement. Similarly, a validated model could be used as an early simulation tool for evaluating new interface concepts before expensive user studies (Oulasvirta et al., 2022). This kind of future normative model would still not necessarily completely replace human testing,

but it could help identify conditions where a display or interaction design is likely to compete with driving-relevant visual sampling.

The framework may also help support human awareness and agency during automated driving transitions. In partially automated vehicles, the driver may need to resume control after a period of reduced visual or cognitive engagement (Braunagel et al., 2017; Zeeb et al., 2016). A model of attention as belief-based information sampling could help estimate whether the driver has an adequate internal representation of the current driving situation before takeover. This would move takeover assessment beyond simple measures of gaze direction or hands-on-wheel status toward a more functional estimate of situation awareness as task-relevant uncertainty. More broadly, this perspective is compatible with human-centered and human-engaged approaches to automation that seek to maintain meaningful human control and agency rather than treating the driver primarily as a passive fallback monitor (Shneiderman, 2022; Ren & Wang, 2026). Future extensions could therefore investigate how adaptive automation and interface support might help maintain or restore task-relevant awareness while preserving an active role for the driver in vehicle control. However, such applications would require extensive validation in more realistic driving environments and against participant-specific data before the model could be reliably used to estimate actual real-time driver states.

5 CONCLUSIONS

The present hierarchical computational rationality model of bounded-optimal visual sampling in lateral and longitudinal vehicle control contributes to attention research by formalizing inattention as active reliance on internal prediction rather than a mere absence of visual input. It suggests that visual attention in driving can be understood as adaptive sampling of task-relevant information under uncertainty. Continuous visual access to every information source is not always necessary, but the appropriate information must be updated before internal estimates become too inaccurate to support safe control. This perspective connects visual sampling, cognitive prediction and vehicle control within a single computational framework.

The model reproduced the broad human pattern across visual sampling timing, attention allocation and longitudinal performance, with all standardized comparison measures falling within one standard deviation of the corresponding human means. At the same time, the absence of lane deviations showed that the current model overestimated the stability of lateral control. The results support the value of hierarchical computational rationality as a framework for explaining how attention may be allocated between interacting control demands, while also identifying the need for more realistic vehicle control variability, richer representations of uncertainty and participant-specific modeling.

In a broader sense, the model demonstrates how adaptive periods of inattention can emerge from the interaction between internal beliefs, prediction error, perceptual costs and task constraints without directly encoding the observed human attention patterns. With further validation and development, the framework could support more context-sensitive models of driver attention that consider not only

observable attention behavior but also the adequacy of task-relevant internal representations, and inform future driver monitoring systems, in-vehicle user interface design and the development of automated driving applications.

Declarations

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval

The study was approved by the Human Sciences Ethics Committee of the University of Jyväskylä (Decision 1129/13.00.04.00/2021) and was conducted in accordance with the University of Jyväskylä's ethical guidelines.

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Use of Generative AI Tools

During manuscript preparation, ChatGPT (OpenAI) and DeepSeek were used to assist with proofreading, language editing, text condensation, manuscript organization and wording refinement for readability. All AI-assisted content was critically reviewed and edited by the authors, who take full responsibility for the final manuscript. These tools were not used to generate, modify or analyze the study data.

Funding

This work was partly supported by the Research Council of Finland (formerly Academy of Finland) through the Appropriate Uncertainty in Manual and Automated Driving project (Grant No. 343259).

Author Contribution

T.H. contributed to conceptualization, methodology, software, formal analysis, visualization, writing the original draft, and reviewing and editing the manuscript. T.K. contributed to conceptualization, methodology, supervision, project administration, funding acquisition, and reviewing and editing the manuscript. T.H.W. contributed to conceptualization, methodology, supervision, and reviewing and

editing the manuscript. X.R. contributed to reviewing and editing the manuscript. All authors approved the final manuscript.

Data Availability

The data supporting the findings of this study are publicly available in https://github.com/drive-in-lab/RCF-JYU-APU-2021-2025/tree/main/Hautaoja%20et%20al%202026%20TBS. The model source code is available at https://gitlab.com/toninikolaialeksanteri/driver-adaptability.

References

1. Abbeel, P., & Ng, A. Y. (2004). Apprenticeship learning via inverse reinforcement learning. Proceedings of the Twenty-First International Conference on Machine Learning (pp. 1–8). Association for Computing Machinery. https://doi.org/10.1145/1015330.1015430
2. Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. Proceedings of the 26th Annual International Conference on Machine Learning (pp. 41–48). Association for Computing Machinery. https://doi.org/10.1145/1553374.1553380
3. Braunagel, C., Rosenstiel, W., & Kasneci, E. (2017). Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness. *IEEE Intelligent Transportation Systems Magazine*, *9*(4), 10–22. https://doi.org/10.1109/MITS.2017.2743165
4. Chandramouli, S., Shi, D., Putkonen, A., De Peuter, S., Zhang, S., Jokinen, J., Howes, A., & Oulasvirta, A. (2024). A workflow for building computationally rational models of human behavior. *Computational Brain & Behavior*, *7*, 399–419. https://doi.org/10.1007/s42113-024-00208-6
5. Chatzikomis, C. I., & Spentzas, K. N. (2009). A path-following driver model with longitudinal and lateral control of vehicle's motion. *Forschung im Ingenieurwesen*, *73*(4), 257–266. https://doi.org/10.1007/s10010-009-0112-5
6. Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477
7. Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–138. https://doi.org/10.1038/nrn2787
8. García, J., & Fernández, F. (2015). A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, *16*(42), 1437–1480.
9. Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273–278. https://doi.org/10.1126/science.aac6076
10. Godthelp, H. (1986). Vehicle Control During Curve Driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *28*(2), 211–221. https://doi.org/10.1177/001872088602800209

11. Godthelp, H., Milgram, P., & Blaauw, G. J. (1984). The development of a time-related measure to describe driving strategy. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(3), 257–268. https://doi.org/10.1177/001872088402600302
12. Grahn, H., Kujala, T., Hautaoja, T., & Salvucci, D. D. (2024). Investigating the situational dynamics of visual information sampling in lateral vehicle control – Subjective vs. objective estimates of spare visual capacity. *Transportation Research Part F: Traffic Psychology and Behaviour*, *107*, 98–114. https://doi.org/10.1016/j.trf.2024.08.034
13. Grahn, H., Kujala, T., Taipalus, T., Lee, J., & Lee, J. D. (2023). On the relationship between occlusion times and in-car glance durations in simulated driving. *Accident Analysis & Prevention*, *182*, 106955. https://doi.org/10.1016/j.aap.2023.106955
14. Guo, K., Ding, H., Zhang, J., Lu, J., & Wang, R. (2004). Development of a longitudinal and lateral driver model for autonomous vehicle control. *International Journal of Vehicle Design*, *36*(1), 50–65. https://doi.org/10.1504/IJVD.2004.005320
15. Hautaoja, T., & Ergonomics (2026). 1–18. Advance online publication. https://doi.org/10.1080/00140139.2026.2654083
16. Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., & Meger, D. (2018). Deep reinforcement learning that matters. Proceedings of the AAAI Conference on Artificial Intelligence, 32(1), 3207–3214. https://doi.org/10.1609/aaai.v32i1.11694
17. Horrey, W. J., Wickens, C. D., & Consalus, K. P. (2006). Modeling drivers' visual attention allocation while interacting with in-vehicle technologies. *Journal of Experimental Psychology: Applied*, *12*(2), 67–78. https://doi.org/10.1037/1076-898X.12.2.67
18. Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review*, *116*(4), 717–751. https://doi.org/10.1037/a0017187
19. Jokinen, J. P. P., Kujala, T., & Oulasvirta, A. (2021). Multitasking in driving as optimal adaptation under uncertainty. *Human Factors*, *63*(8), 1324–1341. https://doi.org/10.1177/0018720820927687
20. Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1–2), 99–134. https://doi.org/10.1016/S0004-3702(98)00023-X
21. Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A. M., & Koch, I. (2010). Control and interference in task switching: A review. *Psychological Bulletin*, *136*(5), 849–874. https://doi.org/10.1037/a0019842
22. Klink, P., Yang, H., D'Eramo, C., Peters, J., & Pajarinen, J. (2022). Curriculum reinforcement learning via constrained optimal transport. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, & S. Sabato (Eds.), Proceedings of the 39th International Conference on Machine Learning (Vol. 162, pp. 11341–11358). PMLR. https://proceedings.mlr.press/v162/klink22a.html
23. Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007

24. Kong, J., Pfeiffer, M., Schildbach, G., & Borrelli, F. (2015). Kinematic and dynamic vehicle models for autonomous driving control design. *2015 IEEE Intelligent Vehicles Symposium (IV)*, 1094–1099. https://doi.org/10.1109/IVS.2015.7225830
25. Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326. https://doi.org/10.1016/j.tics.2006.05.003
26. Kujala, T., Kircher, K., & Ahlström, C. (2023). A review of occlusion as a tool to assess attentional demand in driving. *Human Factors*, *65*(5), 792–808. https://doi.org/10.1177/00187208211010953
27. Kujala, T., & Lappi, O. (2021). Inattention and uncertainty in the predictive brain. *Frontiers in Neuroergonomics*, *2*., Article 718699. https://doi.org/10.3389/fnrgo.2021.718699
28. Kujala, T., & Sarkar, A. (2024). Spare visual capacity and driver inattention in dynamic car following scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour*, *106*, 629–642. https://doi.org/10.1016/j.trf.2024.06.017
29. Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*(2), 279–311. https://doi.org/10.1111/tops.12086
30. Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140. https://doi.org/10.1016/S1364-6613(03)00028-7
31. Oulasvirta, A., Jokinen, J. P. P., & Howes, A. (2022). Computational rationality as a theory of interaction. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, Article 359, 1–14. Association for Computing Machinery. https://doi.org/10.1145/3491102.3517739
32. Pekkanen, J., Lappi, O., Rinkkala, P., Tuhkanen, S., Frantsi, R., & Summala, H. (2018). A computational model for driver's cognitive state, visual perception and intermittent attention in a distracted car following task. *Royal Society Open Science*, *5*(9). https://doi.org/10.1098/rsos.180194. Article 180194.
33. Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M., & Dormann, N. (2021). Stable-Baselines3: Reliable reinforcement learning implementations. *Journal of Machine Learning Research*, *22*(268), 1–8.
34. Regan, M. A., & Oviedo-Trespalacios, O. (2022). Driver distraction: Mechanisms, evidence, prevention, and mitigation. In K. Edvardsson Björnberg, M.-Å. Belin, S. O. Hansson, & C. Tingvall (Eds.), The Vision Zero handbook (pp. 1–62). Springer. https://doi.org/10.1007/978-3-030-23176-7_38-1
35. Ren, X. (2016). Rethinking the relationship between humans and computers. *Computer*, *49*(8), 104–108. https://doi.org/10.1109/MC.2016.253
36. Ren, X., Silpasuwanchai, C., & Cahill, J. (2019). Human-engaged computing: The future of human–computer interaction. *CCF Transactions on Pervasive Computing and Interaction*, *1*(1), 47–68. https://doi.org/10.1007/s42486-019-00007-0
37. Ren, X., & Wang, C. (2026). *Human-Engaged Computing*. Chapman and Hall/CRC.

38. Salvucci, D. D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors*, *48*(2), 362–380. https://doi.org/10.1518/001872006777724417
39. Salvucci, D. D., & Gray, R. (2004). A two-point visual control model of steering. *Perception*, *33*(10), 1233–1248. https://doi.org/10.1068/p5343
40. Sarkar, A., & Kujala, T. (2025). Spare visual capacity and driver inattention in lateral vehicle control. *Transportation Research Part F: Traffic Psychology and Behaviour*, *109*, 1246–1256. https://doi.org/10.1016/j.trf.2025.01.043
41. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms [Preprint]. arXiv. https://arxiv.org/abs/1707.06347
42. Senders, J. W., Kristofferson, A. B., Levison, W. H., Dietrich, C. W., & Ward, J. L. (1967). The attentional demand of automobile driving. *Highway Research Record*, *195*, 15–33.
43. Shneiderman, B. (2022). Human-Centered AI. Oxford University Press. https://doi.org/10.1093/oso/9780192845290.001.0001
44. Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99–118. https://doi.org/10.2307/1884852
45. Summala, H., Nieminen, T., & Punto, M. (1996). Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors*, *38*(3), 442–451. https://doi.org/10.1518/001872096778701944
46. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
47. van Winsum, W., & Godthelp, H. (1996). Speed choice and steering behavior in curve driving. *Human Factors*, *38*(3), 434–441. https://doi.org/10.1518/001872096778701926
48. Wilson, R. C., & Collins, A. G. E. (2019). Ten simple rules for the computational modeling of behavioral data. *eLife*, *8*, e49547. https://doi.org/10.7554/eLife.49547
49. Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis & Prevention*, *92*, 230–239. https://doi.org/10.1016/j.aap.2016.04.002

Figures



Figure 1

University of Jyväskylä driving simulator. During the combined task, the speedometer visible on the center monitor was hidden from the road view and presented on the rightmost monitor (see Fig. 2).



Figure 2

Simulator task view during a single 0.5 s unoccluded visual sample.

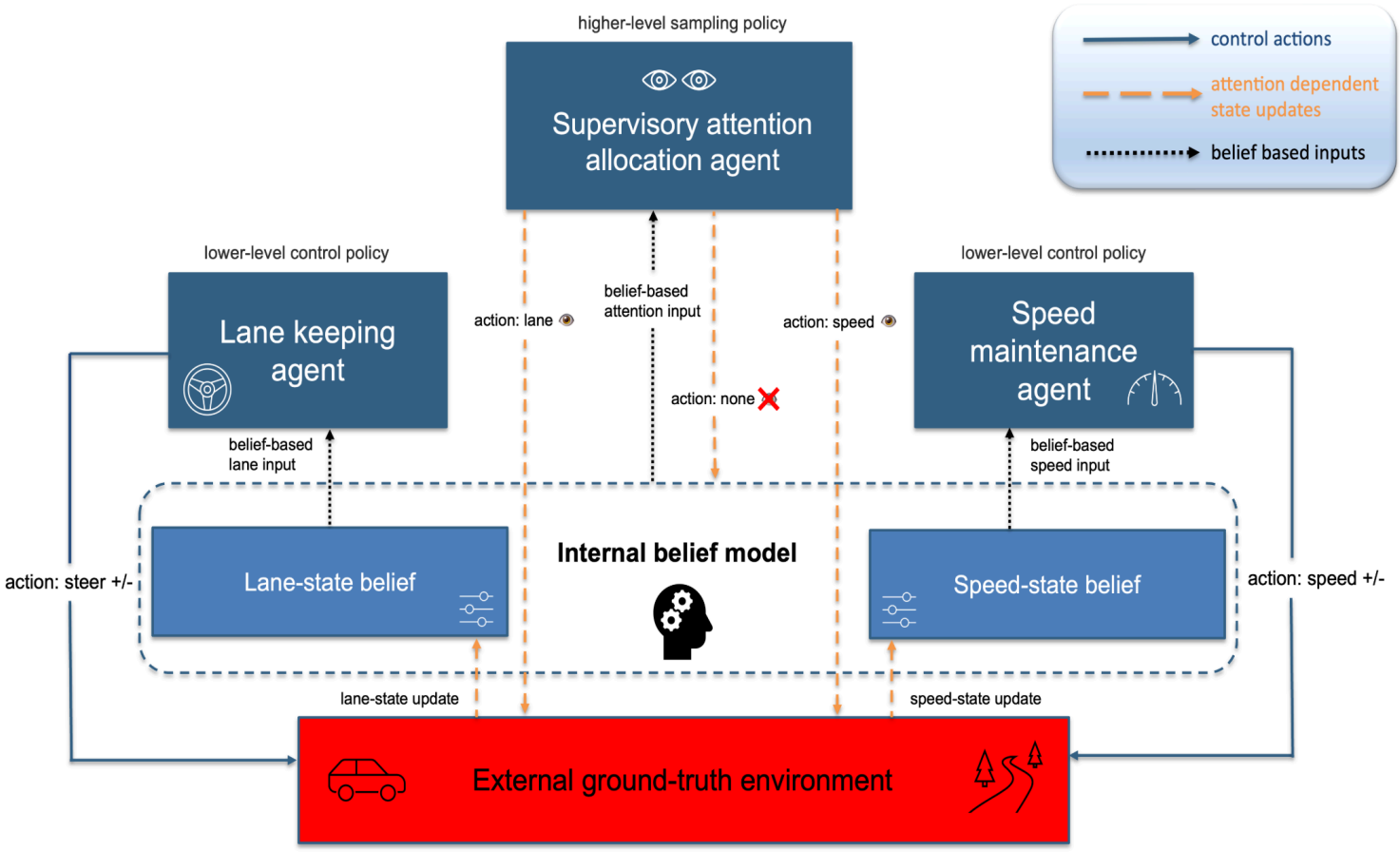


Figure 3

Structure of the hierarchical computational rationality model. Separate lower-level policies control lateral steering and longitudinal speed, while a supervisory attention allocation policy selects whether to sample lane-channel, speed-channel or remain occluded. An internal attention-dependent belief model maintains estimates of the hidden lane and speed states, while the external environment represents the true vehicle and road state.

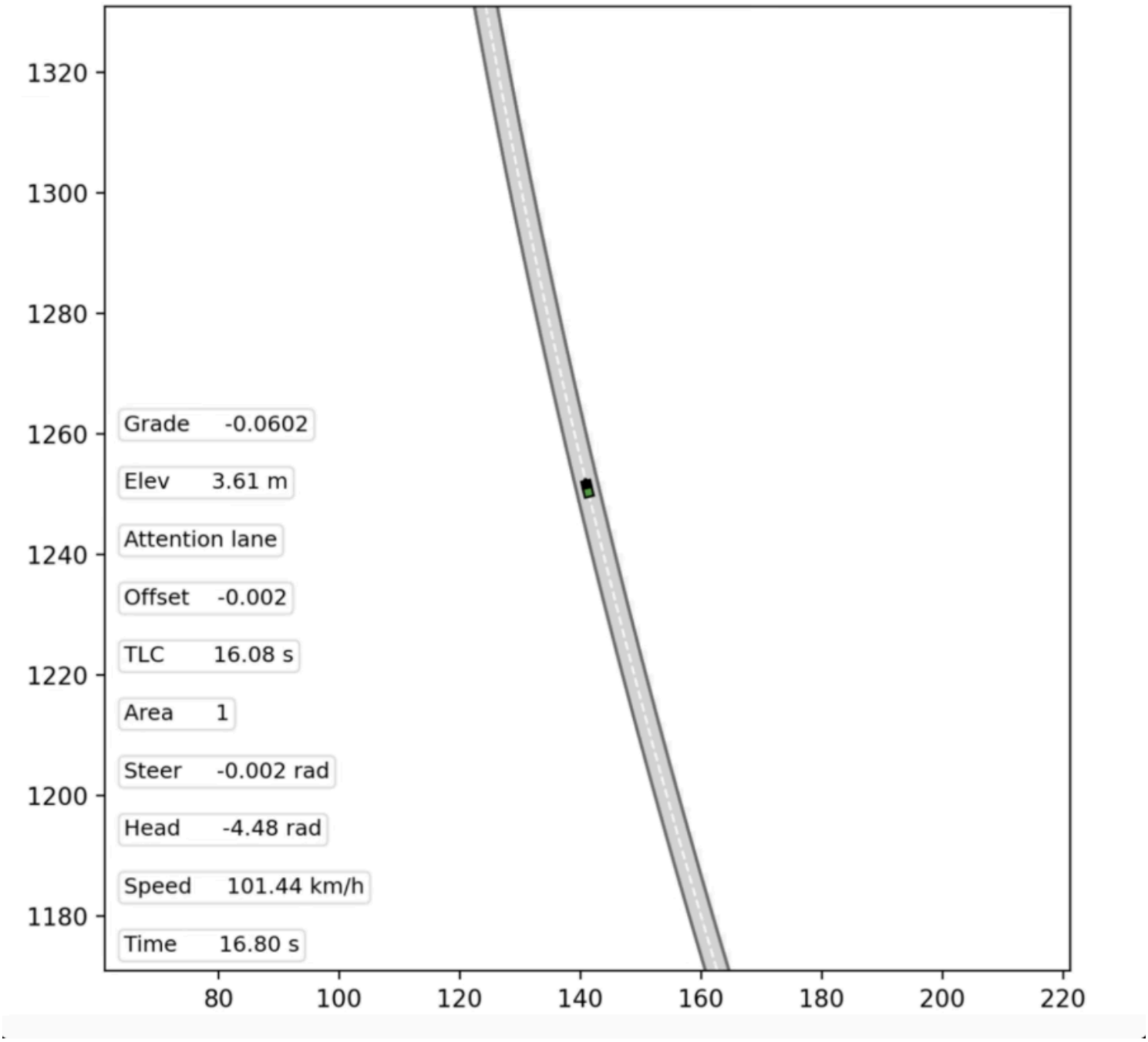


Figure 4

Example frame from the model's animated simulation view during an evaluation episode.

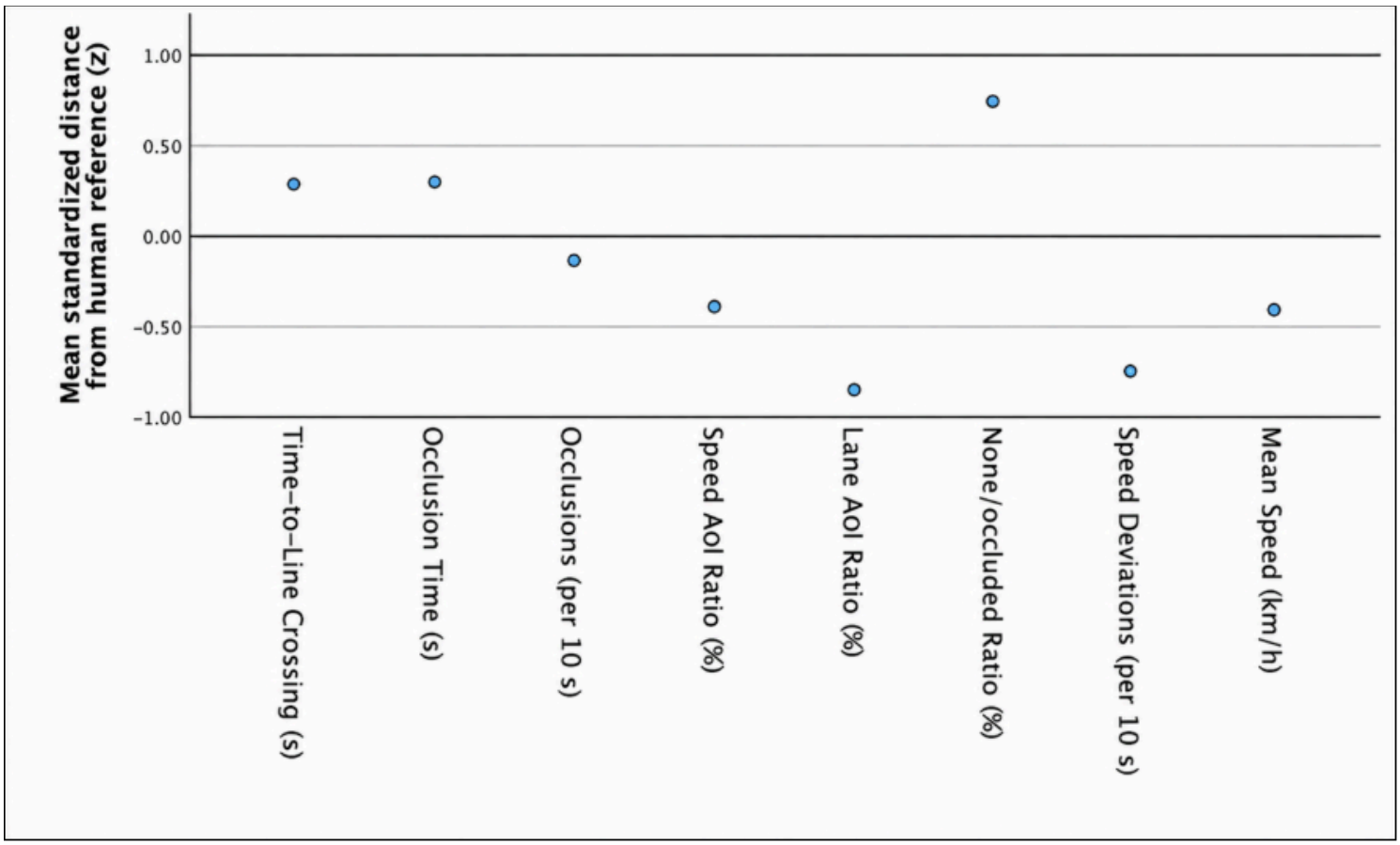


Figure 5

Standardized distances between the model and human means across the eight evaluation measures.

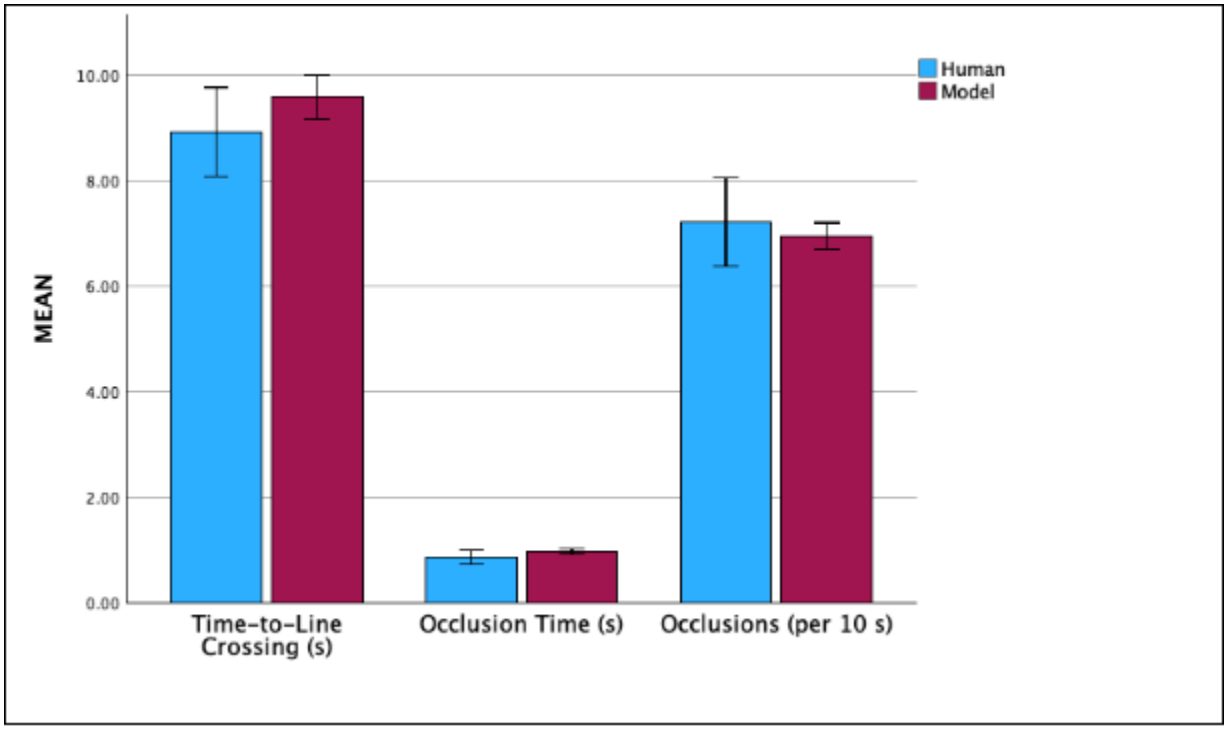


Figure 6

Comparison of human and model visual sampling measures. The panels show mean OT, mean TLC (approximation at occlusion onset) and the number of occlusions per 10 s. Error bars represent 95% confidence intervals calculated across participant means for the human data and across stochastic run means for the model data. Model intervals reflect variation across repeated runs of the same policy, not between-driver variability.

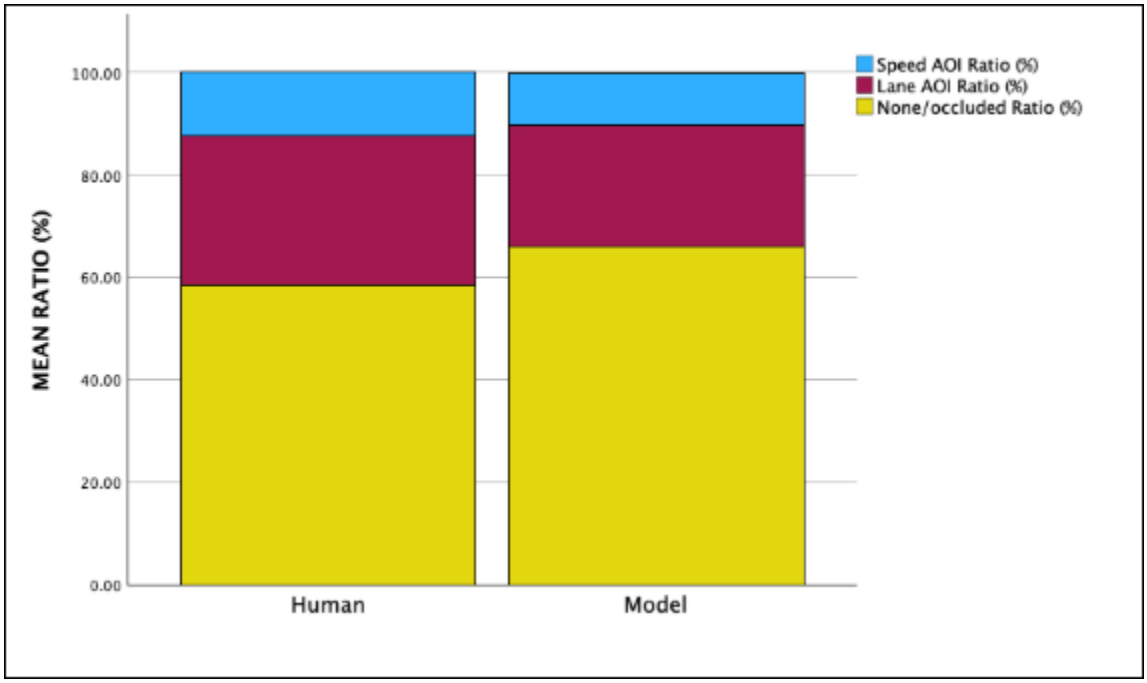


Figure 7

Mean allocation of time to the Lane, Speed and None categories in the human and model data. Values are presented as 100% stacked bars showing the relative distribution of attention across the three categories. The human results are based on 27 participants.

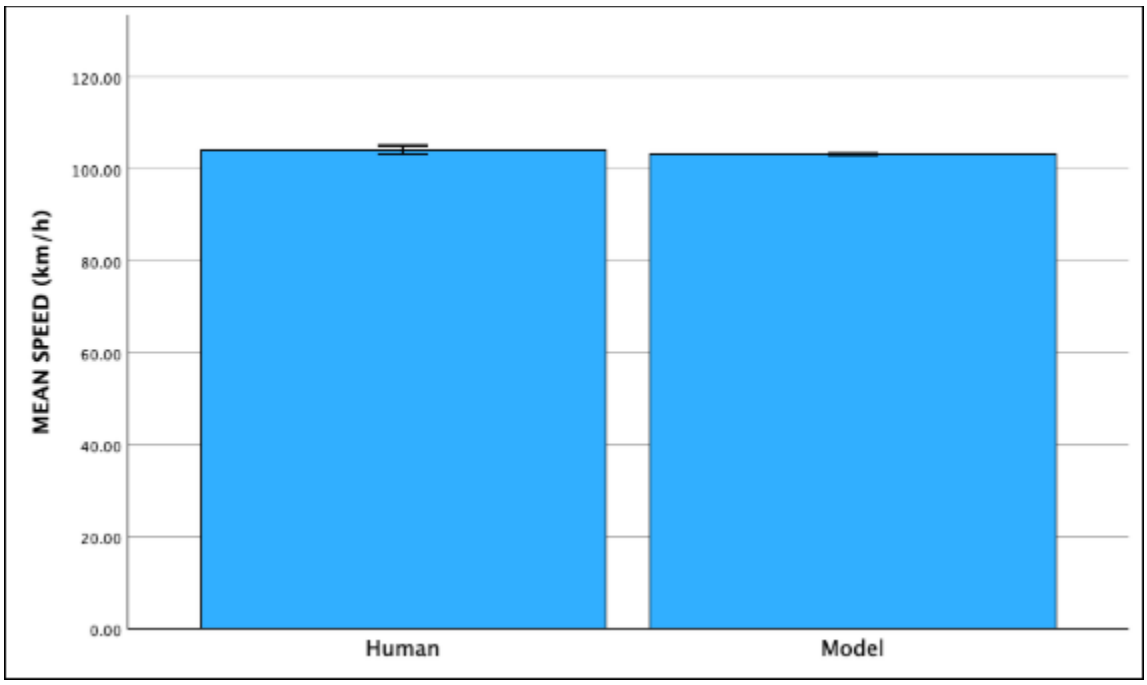


Figure 8

Comparison of human and model mean driving speed. Error bars represent 95% confidence intervals across human participant means and model simulation-run means.

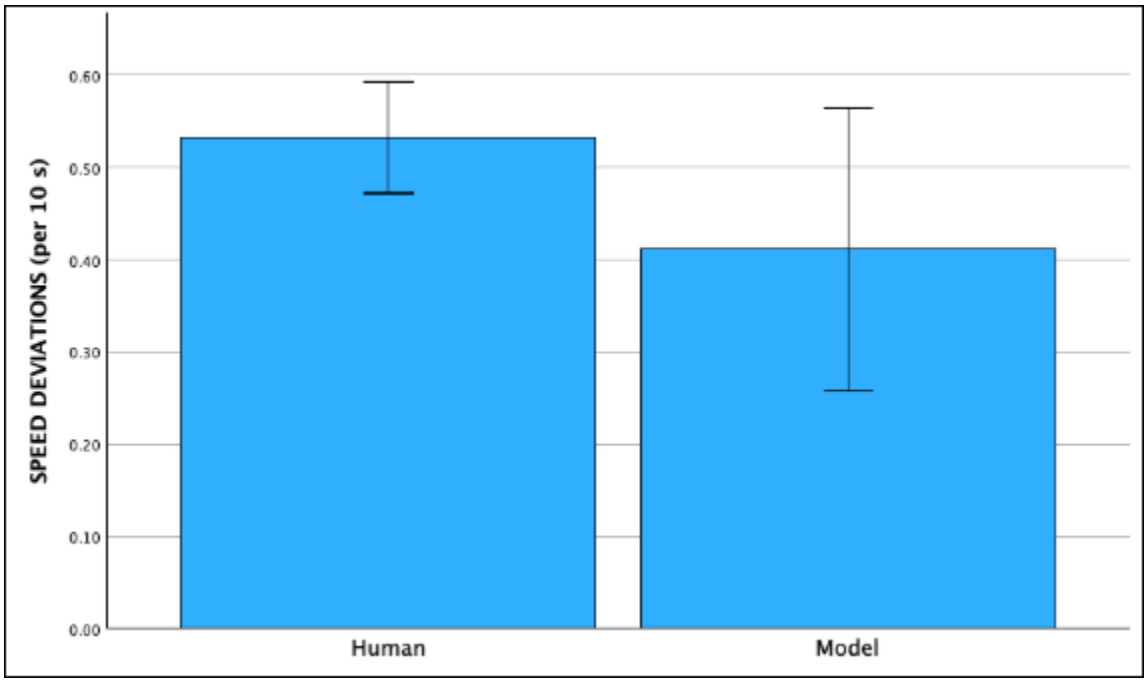


Figure 9

Comparison of human and model speed deviation rates per 10 s. Error bars represent 95% confidence intervals across human participant means and model simulation-run means.

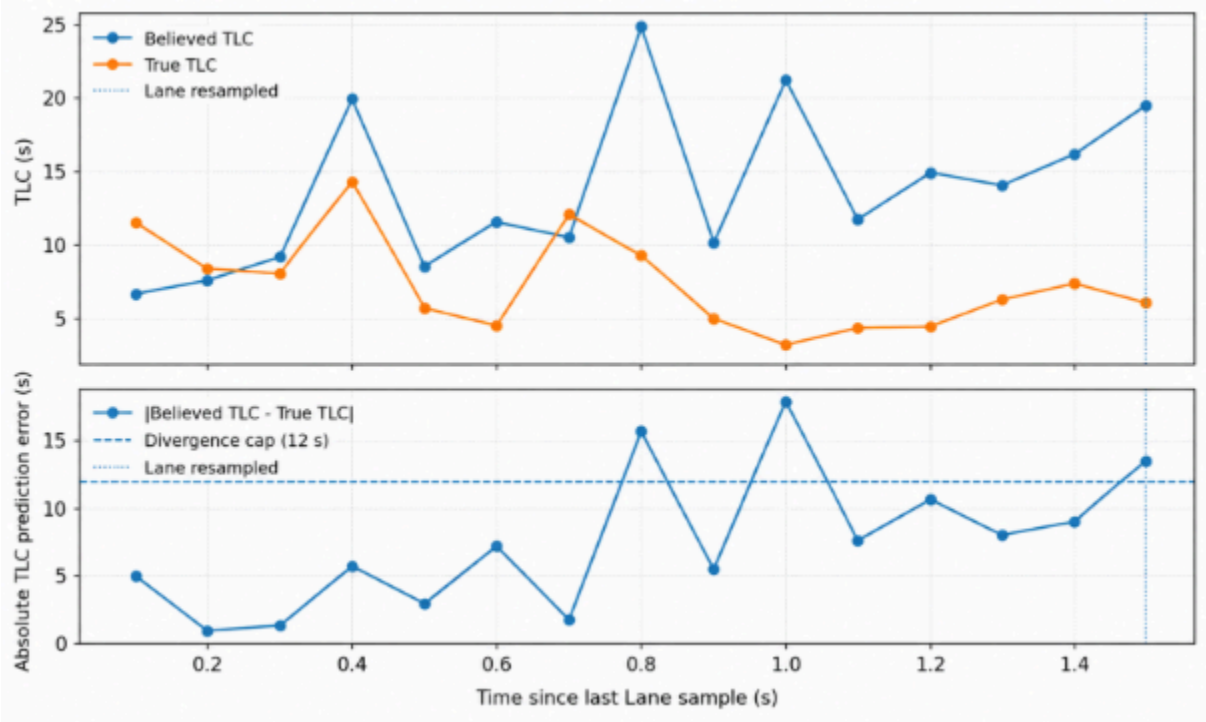


Figure 10

TLC belief–truth prediction error during an illustrative lane-occlusion episode. The upper panel shows believed and true TLC, while the lower panel shows their absolute difference. The vertical dotted line indicates the next Lane sample, with the final value representing the pre-synchronization prediction error. The horizontal dashed line indicates the 12 s cap used in the lane-divergence component of the attention reward.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- APPENDIX.docx