Contents lists available at ScienceDirect

International Journal of Human - Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Effects of in-vehicle auditory interactions on takeover performance in SAE L2 semi-automated vehicles

Jiwoo Hwang, Woohyeok Choi *, Auk Kim *

Kangwon National University, 1 Gangwondaehakgil, Chuncheon 24341, South Korea

ARTICLE INFO

Keywords:
Automated vehicle
Auditory secondary task
Physiological sensing
Driver distraction

ABSTRACT

Commercially available automated vehicles require drivers to maintain focus on their driving environment and be prepared to fully control their vehicles (i.e., perform a takeover) when critical incidents occur (e.g., sudden automation failures). Therefore, drivers are discouraged from engaging in non-driving tasks that cause visual or manual distractions. Auditory interactions, despite being considered a safe alternative, can consume the attentional resources of drivers, causing them to respond poorly to critical situations. This study investigates (1) how varying levels of auditory interactions affect takeover performance and (2) what physiological contexts are related to the takeover performance in SAE Level 2 automated driving. For the investigation, 50 drivers wore wearable devices that collected various physiological signals and performed six different auditory tasks during L2 automated driving in a simulator-based experiment. The results showed that auditory interactions could degrade the takeover performance and that the task demand for auditory interactions nonlinearly affected the takeover performance, possibly owing to behavior changes intended to prevent the task difficulty from becoming excessively high. Additionally, physiological contexts such as pupil diameter, dispersion of eye movements, and interbeat interval, were found to be related to the takeover performance. Subsequently, we discussed drivers' behavior changes, practical deployment of in-situ physiological measures, and design implications for mitigating the degradation of takeover performance due to auditory tasks.

1. Introduction

Commercially available automated vehicles can perform many driving-related tasks. However, human involvement is still crucial in automated driving. For example, although the automated driving systems embedded in current commercial vehicles (level 2 or partially automated driving systems) can manage the longitudinal acceleration and lateral position of a vehicle (e.g., adaptive cruise control and lane-keeping systems), driver intervention is still required when automation fails (SAE International, 2021). In other words, although manual operation of the steering wheel and pedals is not necessary, drivers are still required to perform an *Object and Event Detection and Response* (OEDR) task (SAE International, 2021, Council of European Union 2022), which involves consistently monitoring driving situations and intervening (i.e., takeover; taking over the control as soon as automated driving fails). Even if the next advanced automated system (level 3 or conditionally automated driving system) becomes commercially available, it can only be activated in limited circumstances, such as roads that are exclusive to heavy vehicles (e.g., no pedestrians, bicycles, and motorcycles) (UNECE, 2021). Therefore, the L2 system is expected to be utilized in most situations.

Although the current automated driving system requires drivers to perform an OEDR task, studies have indicated that during automated driving, drivers often engage in secondary tasks, which can cause driver distractions, diverting their attention from the driving environment. Specifically, secondary tasks refer to any non–driving related activity that drivers perform while driving (or during automated driving) (Regan et al., 2008). An observational study by Banks et al. (2018) showed that during automated driving (e.g., Tesla Autopilot mode), drivers often took their hands off the steering wheel and engaged in visual–manual secondary tasks (e.g., drinking coffee), possibly due to their complacency and trust in automated systems. Additionally, drivers are more likely to stop looking ahead and perform visual–manual secondary tasks during automated driving than during manual driving (Naujoks et al., 2016, Solis-Marcos et al., 2018). Engaging in visual–maual secondary tasks during automated driving can cause visual distractions, which reduce the ability of a driver to respond if the automation system fails, resulting in various risky situations. For example, visual-manual

* Corresponding authors.
E-mail addresses: hwangjiwoo@kangwon.ac.kr (J. Hwang), woohyeok.choi@kangwon.ac.kr (W. Choi), kimauk@kangwon.ac.kr (A. Kim).

https://doi.org/10.1016/j.ijhcs.2024.103401
Received 11 November 2023; Received in revised form 28 October 2024; Accepted 28 October 2024

interactions during automated driving cause drivers to take longer to resume manual driving and, after the resumption of control, degrade their lateral control (Louw et al., 2019).

As an alternative to visual–manual interactions, auditory interfaces are often a preferred option for providing in-vehicle interactions (Ayoub et al., 2019, Alvarez et al., 2015, Sodnik et al., 2008). Secondary tasks involving auditory interactions do not follow the same modalities as those of monitoring tasks, which require the visual–manual operations of the driver. Therefore, while engaging in auditory tasks, drivers can simultaneously perform an OEDR task. However, literature hints that auditory tasks can cause cognitive distractions, which can potentially degrade the performance of monitoring tasks, leading to a worse takeover performance. For example, multiple-resource theory (Wickens, 2008) implies that drivers have limited attentional (or cognitive) resources. Therefore, the performance in either or both of driving and secondary tasks can decrease if the overall cognitive resources required for the tasks exceed the driver's limited cognitive resources. Several studies have already shown that in the manual-driving context, auditory tasks can lead to distracted driving (Kim et al., 2020, Strayer et al., 2015, Strayer et al., 2017, Loew et al., 2023).

While in-vehicle auditory tasks are conjectured to distract drivers and thus degrade their ability to counteract automation failures, to the best of our knowledge, the effects of auditory interactions in current L2 automated vehicle contexts are still under-investigated. Therefore, in this study, we investigated (1) how varying levels of auditory secondary tasks affect the takeover performance in L2 driving contexts and (2) which are the typical physiological contexts related to the takeover performance in L2 driving contexts. Our investigation would lay a foundation for developing an intelligent interruption management system that automatically estimates whether a driver can safely engage in auditory tasks during L2 automated driving by monitoring the driver with physiological sensors.

2. Background and related works

2.1. Driving automation levels and roles of drivers

As shown in Table 1, According to SAE (SAE International, 2021), the levels of driving automation are classified into six levels from Level 0 (L0, fully manual) to Level 5 (L5, fully automated). Depending on the automation level and the aspects of the automated driving task, drivers need to perform different sets of the following driving tasks: (1) steering and acceleration, wherein drivers control the lateral (e.g., steering wheel) and longitudinal (e.g., acceleration and deceleration) positions of their vehicles; (2) monitoring driving situations, wherein drivers constantly perceive, understand, and predict objects and events on a road; and (3) executing a fallback procedure when the automation system fails, wherein drivers take over the control of a vehicle from the system as a response in critical situations where automation does not operate appropriately.

Except for L0 (i.e., a fully manual driving level with no automation), driving tasks are supported (partially) by automated driving systems. Regarding systems at levels L1 - L4, they can only operate under certain conditions. In an L1 system, the automation system controls either the lateral or longitudinal position of the vehicle, and the driver needs to control the other position. Vehicles with the L2 system can manage both the longitudinal and lateral positions. Although drivers on L2 driving are not required to operate the steering wheel and pedals, they need to perform OEDR tasks (i.e., monitoring the driving environment and resuming control in case the automation fails). In an L3 system, the vehicle controls both the longitudinal and lateral positions and always initiates a takeover request when the driver needs to take over the control; therefore, the drivers no longer need to monitor driving situations. In an L4 system, under limited circumstances, the system supports fully automated driving and does not require takeover. Finally, in an L5 system, the system fully supports automated driving under all conditions.

This study focuses on L2 driving automation, which is the automation level of most current commercial automated vehicles (e.g., Tesla Autopilot). As previously mentioned, at this level, the drivers need to engage in the OEDR tasks although they do not need to manually control their vehicles while automated driving is activated. Until L5 (or fully automated) vehicles are commercially available, L2 systems are likely to be used in most circumstances because UN Regulation No. 157 restricts the operation of advanced systems to roads that are physically separated from opposite traffic, with no pedestrians or cyclists (UNECE, 2021).

2.2. Effects of secondary tasks on takeover performance in semi-automated vehicles

As shown in Table 2, prior studies on the effects of secondary tasks on takeover performance mainly considered L3 systems, where the drivers need to resume control of their vehicle as soon as possible after a takeover request is triggered. It is well known that a time delay occurs while individuals are switching from one to another task (Janssen et al., 2019). Thus, studies involving L3 system have investigated the time taken by a driver to switch from a visual–manual secondary task to a takeover task in response to a takeover request. For example, Mok et al. (Mok et al., 2017) investigated the time budget required for a driver playing a game on a tablet PC to successfully take over in response to a takeover request. While these studies provided important insights, their findings cannot be directly applied in L2 contexts. In L2 contexts, drivers need to simultaneously perform monitoring and secondary tasks, which is different from L3 contexts, wherein such multitasking is unnecessary.

In L2 driving contexts, auditory interactions are more favorable for technology usage than visual–manual interactions, considering that auditory interactions do not involve the visual and manual operations present in the OEDR task (i.e., drivers can perform monitoring and auditory tasks simultaneously). However, only few studies have focused on the effects of auditory tasks on takeover performance. Notably, previous studies have not considered the demands of different auditory tasks. For example, Blommer et al. (2015) examined how the takeover performance varied between two conditions where drivers performed a visual–manual task (i.e., watching a video) and an auditory task (i.e., listening to the radio) while simultaneously performing a monitoring task.

Table 1
SAE levels of driving automation (SAE International, 2021). O = required/supported, △ = partially required, X = not required/not supported.

Level	Condition for Driver's task during automation System's supporting automation	Steering & acceleration	Monitoring driving situations	Fallback when the system fails	takeover request
0: no driving automation	None	O	O	O	X
1: driver assistant	Limited	△	O	O	X
2: partial driving automation	Limited	X	O	O	X
3: conditional driving automation	Limited	X	X	O	O
4: high driving automation	Limited	X	X	X	X
5: full driving automation	Always	X	X	X	X

Table 2
Summary of secondary task types in previous research.[1] SURT (Surrogate reference task): searching for a circle that differs in size from others.[2] Describing images: with three images and scrambled letters, completing a word related to every image by combining the letters. St. task = standardized task; Natural. task = Naturalistic task. (References (Liu et al., 2024; Du et al., 2024; Sanghavi et al., 2023; Politis et al., 2017; Wang et al., 2022; Chen and Chen, 2021; Pakdamanian et al., 2021; Morando et al., 2021; Du et al., 2020a; Du et al., 2020b; Eriksson et al., 2019; Berghöfer et al., 2018; Wandtner et al., 2018; Wan and Wu, 2018; van der Heiden et al., 2017; Feldhütter et al., 2017; Mok et al., 2017; Mok et al., 2015; Bueno et al., 2016; Gold et al., 2016; Gold et al., 2015; Walch et al., 2015; Schwalk et al., 2015; Radlmayr et al., 2014; Louw et al., 2019; Blommer et al., 2015; Arkonac et al., 2019; Yang et al., 2021) is cited in Table body part).

Studies	Level	Visual tasks																			Auditory tasks						
		Standardized task								Naturalistic task											St. task		Natural. task				
		1-back task	2-back task	SuRT[1]	Describing images[2]	Reading road signs	Selecting a shape	Reading & typing	Solving arithmetic	Game	Email	Texting	Web browsing	Reading	Watching a video	Video transcription	Calendar task	Backseat searching	Temperature control	Music selection	2-back task	Repeating verbally	Radio	Audiobook	Conversation with human	Conversation with agent	Game
Liu et al., 2024	3											O		O													
N. Du et al., 2024	3		O																								
Sanghavi et al., 2023	3									O																	
Politis et al., 2023	3									O																	
Wang et al., 2022	3																									O	
Chen and Chen, 2021	3													O													
Pakdamanial et al., 2021	3								O			O	O	O											O		
Morando et al., 2021	3									O																	
N. Du et al., 2020a	3	O	O																								
N. Du et al., 2020b	3	O	O																								
Erikson et al., 2019	3									O																	
Berghöfer et al., 2018	3									O				O				O						O			
Wandtner et al., 2018	3							O														O					
Wan et al., 2018	3							O		O				O	O												
Heiden et al., 2017	3															O	O										
Feldütter et al., 2017	3			O																							
Mok et al., 2017	3									O																	
Mok et al., 2015	3														O												
Bueno et al., 2016	3				O																						
Gold et al., 2016	3																										O
Gold et al., 2015	3			O							O										O						
Walch et al., 2015	3														O												
Schwalk et al., 2015	3									O				O													
Radlmayr et al., 2014	3			O																	O						
Louw et al., 2019	2					O	O																				
Blommer et al., 2015	2														O								O				
Arkonac et al., 2019	2														O												
Yang et al., 2021	2										O								O	O							

Multiple Resource Theory (Wickens, 2002, Wickens, 2008) indicates that in an L2 automated driving context, auditory tasks can potentially degrade the takeover performance. According to Multiple Resource Theory, human cognitive resources are divided into multiple channels or pools, each associated with different modalities (e.g., visual or auditory). When two tasks utilize different resources or modalities (e.g., performing visual and auditory tasks simultaneously), the tasks are less likely to interfere with each other, as they draw from separate cognitive pools. However, despite utilizing different modalities, the performance of these tasks can still decrease. This decline is due to the overall limited capacity of human cognitive resources (i.e., human central processor) (Wickens, 2002, Wickens, 2008, Moray, 1967). In other words, when the combined cognitive demand of the two tasks exceeds this total capacity, the performance of the tasks can decrease due to the insufficient available resources to perform both tasks. This implies that, in L2 automated driving contexts, if the cognitive demand of an auditory task is high, drivers are more likely to have insufficient resources to perform both auditory and monitoring tasks. In such cases, the performance of either or both tasks can be degraded depending on the human's resource allocation policy (Wickens, 2008, Navon and Gopher, 1979). The driver may: (1) allocate sufficient resources to monitoring, reducing those available for the auditory task, (2) allocate sufficient resources to the auditory task, reducing those available for monitoring, or (3) allocate insufficient resources to both tasks. In scenarios (2) and (3), the performance of the driver's monitoring task would decrease (Strayer and Johnston, 2001).

Given that a monitoring task requires a considerable amount of mental resources, and auditory tasks are more cognitively demanding than visual– manual interactions (Faure et al., 2016), it is crucial to investigate the performance impact of auditory tasks at various levels of cognitive demand. However, no previous studies have focused on the takeover performance when drivers perform various auditory tasks. Several studies have investigated the impact of auditory interactions on driving in manual driving contexts (Kim et al., 2020, Strayer et al., 2015, Strayer et al., 2017, Loew et al., 2023). For example, Strayer et al. (2015) found that the larger the cognitive load of an auditory secondary task, the slower the braking reaction time of a driver response to the sudden braking of a leading vehicle. In addition to multiple resource theory, these findings also hint that in L2 driving contexts, auditory interactions may degrade the takeover performance. However, these findings may not be directly applicable to L2 driving contexts since the primary task differs between manual driving contexts (i.e., monitoring and maneuvering tasks) and L2 driving contexts (i.e., only a monitoring task).

2.3. In-vehicle auditory secondary tasks

In vehicular contexts, prior studies have considered various auditory secondary tasks. These secondary tasks can be categorized as (1) standardized tasks and (2) naturalistic tasks. Standardized tasks are designed for comparability and repeatability, although they may not fully represent real-world situations. For example, the *n*-back task (or delayed digit recall task) has been widely applied in distracted driving scenarios in both automated driving contexts (Gold et al., 2015, Radlmayr et al., 2014) as well as manual driving contexts (Kim et al., 2018, Kim et al., 2020). It induces systematically structured cognitive demands depending on the number of delayed digits (i.e., *n*) that participants need to recall (Mehler et al., 2011); a larger value of *n* corresponds to greater cognitive demand. In our work, we considered the n-back task as a standardized task and adhered to the standard procedure outlined in (Mehler et al., 2011).

Naturalistic tasks have also been widely considered owing to their high generalizability in real-world scenarios (Pakdamanian et al., 2021, Gold et al., 2016). These naturalistic tasks include both uni-directional interactions (listening only; e.g., listening to a radio (Blommer et al., 2015) or audiobook (Berghöfer et al., 2018)), and bi-directional interactions (listening and speaking; e.g., a conversation with a voice assistant (Wang et al., 2022)). Uni-directional interactions require only language comprehension, whereas bi-directional interactions require both language comprehension and production. Consequently, uni-directional interactions generally impose lower cognitive demand compared to bi-directional interactions (Lee et al., 2017, Rann and Almor, 2022). Furthermore, prior research has found differences in the cognitive demands between arithmetic and linguistic auditory tasks (Horrey et al., 2009). Our work expands upon previous research by considering a wider range of naturalistic tasks with varying cognitive loads, encompassing both uni-directional and bi-directional interactions, as well as arithmetic and linguistic tasks.

2.4. Takeover scenarios and performance metrics

Prior studies on takeover performance in automated driving contexts have considered various takeover scenarios (e.g., jaywalking (van der Heiden et al., 2017, Wang et al., 2022), a sudden stop of a leading vehicle (Wan and Wu, 2018), and a sudden appearance of an obstacle (Pakdamanian et al., 2021, Wan and Wu, 2018)). These scenarios can be classified into two situations depending on whether the vehicle provides a takeover request (i.e., whether it is L2 or L3): (1) a takeover situation without a request (L2), wherein the driver needs to take over the control of a vehicle in response to such event, and (2) a takeover request situation (L3), wherein the driver needs to take over the control of a vehicle in response to a takeover request. L2 vehicles are still limited to offering takeover requests to drivers. Therefore, prior studies on the L2 vehicles have considered takeover situations without a request (Louw et al., 2019, Blommer et al., 2015, Arkonac et al., 2019). For example, Blommer et al. (2015) considered a takeover scenario where a driver needed to take control after a leading vehicle suddenly maneuvered to the left lane, in response to a concealed vehicle that had stopped directly ahead of the leading vehicle. In contrast, prior studies on L3 vehicles have considered takeover request situations since these vehicles always provide takeover requests when a driver needs to take over the control of a vehicle. For example, Wang et al. (2022) considered various takeover request situations in which takeover requests had been provided due to different types of events (e.g., pedestrians jaywalking, encountering road construction ahead, etc.). Since our work was conducted in L2 driving contexts, our scenario did not include the takeover request.

To assess the effect of secondary tasks on the takeover performance, prior studies considered various takeover performance measures. These measures can be divided into two main categories: (1) takeover reaction time and (2) takeover quality. The takeover reaction time represents how quickly a driver reacts to takeover requests or critical events. In L2 contexts, it is defined as the elapsed time from the onset of a critical event (e.g., system malfunction) to the first moment of a reaction (Louw et al., 2019). In contrast, studies on L3 contexts have considered the elapsed time from the onset of a takeover request to the first moment of the reaction (Chen and Chen, 2021, Wan and Wu, 2018). The takeover quality focuses on how well drivers transition from automated to manual driving after the takeover. It has been widely considered to compensate for time-based measures (e.g., takeover reaction time) because the time-based measures are insufficient for indicating the capability of avoiding risk (or the safety levels in a takeover situation). Other factors, such as situational awareness and takeover behavior (e.g., braking, steering, or both) also affect the capability (Louw et al., 2017, Blommer et al., 2017). Therefore, takeover quality is an important measure of takeover performance and has recently attracted the attention of researchers. Different measures are used, depending on the takeover situation. For example, the collision (including driving onto a sidewalk) was measured in a collision-avoidance situation (Wandtner et al., 2018; Wan and Wu, 2018). In a curved or lane-keeping situation, the deviation of the lateral position or steering-wheel angle of a vehicle was measured (Mok et al., 2017; Pakdamanian et al., 2021). Since our work considered collision-avoidance situations in the L2 driving context, we measured takeover reaction time and collision.

2.5. Estimating driver distraction using sensor data

To ensure the safety of drivers in automated vehicles, it is important to estimate driver distraction. Driver distraction refers to drivers diverting their attention from driving to engage in secondary tasks. Prior studies on the automated driving context, researchers mainly considered visual distractions–driver distractions caused by the visual–manual interaction. They leveraged physiological contexts (mostly eye gaze information) of drivers obtained from wearable devices. For examples, Berghöfer et al. (2018) analyzed the off-road glance behavior of drivers in the occurrence of visual distractions in L3 contexts and found that gaze information could be used to estimate the task-switching time (or takeover reaction time). In L2 contexts, Louw et al. (2019) examined the glance behaviors of drivers and observed differences in glance duration across different visual–manual secondary tasks.

Alongside visual–manual tasks, auditory–verbal tasks with large cognitive load also contributes to driver distraction in the context of manual driving (Kim et al., 2020, Strayer et al., 2015, Strayer et al., 2017). In this case, sensor data associated with cognitive load has been used to estimate cognitive distractions, which are the driver distractions caused by the auditory–verbal tasks. Prior studies have found that physiological contexts, such as galvanic skin response (GSR) (Mehler et al., 2009), heart rate (HR) (Mehler et al., 2009), inter-beat interval (IBI) (Henelius et al., 2009), gaze dispersion (Gold et al., 2016), and pupil diameter (Kun et al., 2013), are associated with cognitive load. According to these studies, as cognitive load increases, GSR, pupil diameter, gaze dispersion, and HR increase, while IBI decreases. However, to our knowledge, no studies have focused on how physiological contexts can be leveraged to estimate takeover performance in the context of auditory interactions in L2 driving environments.

3. Research questions and hypotheses

Given the multiple resource theory and empirical studies, in vehicular contexts, auditory interfaces are becoming increasingly preferred (Ayoub et al., 2019, Alvarez et al., 2015, Sodnik et al., 2008, Wickens, 2002). Current L2 automated driving systems allow drivers to engage in auditory interactions, as long as they keep their eyes on the road and/or their hands on the steering wheel (Barry, 2022). However, according to Multiple resource theory (Wickens, 2002, Moray, 1967), auditory secondary tasks can potentially degrade the takeover performance depending on their cognitive demands. To the best of our knowledge, in L2 driving contexts, no study has focused on the takeover performance of drivers when they engage in auditory tasks with varying cognitive demands. In manual driving contexts, several studies revealed these potential hazards when the cognitive demand of auditory tasks was high (Kim et al., 2020, Strayer et al., 2015, Strayer et al., 2017, Loew et al., 2023). Extending this, our first research question is: "How varying levels of auditory secondary tasks affect the takeover performance in L2 driving contexts." We hypothesize that during L2 automated driving, auditory tasks with high cognitive demands will degrade the driver's takeover performance because the drivers have insufficient cognitive resources for both auditory and OEDR tasks. Whereas, auditory tasks with low cognitive demands will not affect takeover performance because the drivers have sufficient cognitive resources for both tasks.

H1a: Auditory tasks with high cognitive demands will degrade the driver's takeover performance.

H1b: Auditory tasks with low cognitive demands will not degrade the driver's takeover performance.

If there is any decline in the takeover performance owing to auditory interactions, it is essential to identify potentially challenging situations wherein drivers may not appropriately take over the control of a vehicle in advance. One feasible approach could be to monitor and analyze the physiological contexts of the drivers. Accordingly, our second research question is: "Which are the typical physiological contexts related to the takeover performance in L2 driving contexts." To identify relevant physiological measurements, we collected wearable sensor data associated with cognitive distractions (e.g., GSR, HR, IBI, pupil diameters, gaze dispersion, and off-road glance rate), as well as data related to visual distraction (e.g., off-road glance rate). We hypothesize that when drivers engaged in auditory secondary task during L2 driving, physiological contexts associated with cognitive distractions can be used to estimate takeover performance. Conversely, since drivers can engage in auditory interaction without visual distractions, physiological contexts associated with visual distraction may not be effective for estimating takeover performance in L2 driving.

H2a: The physiological contexts related to cognitive distractions, such as GSR, HR, IBI, pupil diameters and gaze dispersion, will differ depending on the takeover performance of a driver performing auditory tasks.

H2b: The physiological contexts related to visual distractions, such as off-road glance rate, will not differ depending on the takeover performance of a driver performing auditory tasks.

4. Methods

4.1. Study design

In this study, we first selected six auditory secondary tasks and measured their task-demand level in Section 4.4. Next, we conducted a within-subject design experimental study by simulating an L2 automated vehicle in which the drivers performed the six auditory tasks while simultaneously performing an OEDR task. Details of the experimental procedure and an OEDR task are described in Section 4.3 and Section 4.7. During the experiment, we measured the takeover reaction times and success rates to answer RQ1 and collected the drivers' physiological signals to answer RQ2.

4.2. Apparatus

4.2.1. Automated vehicle simulator

We conducted an experiment using an automated vehicle simulator with a vehicle cockpit module, as shown in Fig. 1. We considered a driving simulator for driver safety as our experiment involved takeover scenarios (details are illustrated in Section 4.3) wherein any failure in takeover could lead to unfavorable outcomes (e.g., collision with a leading vehicle). The module consisted of a steering wheel, an accelerator, and a brake pedal. We developed automated driving simulation software based on CARLA (Dosovitskiy et al., 2017)—an open-source platform for developing and testing automated driving algorithms—to fully support a takeover scenario with the vehicle cockpit module. The control of the vehicle was instantly returned to the driver (i.e., a transition from automated driving to manual driving mode) when the driver either (1) manually operated the steering wheel and pedals or (2) pressed a button on the steering wheel. In addition, the simulator logged the vehicle control information (e.g., acceleration, brake, and steering), position, and direction at 20 Hz.

Fig. 1. L2 automated vehicle simulator with a cockpit module: (1) a steering wheel; (2) a brake pedal; (3) a speakerphone.

4.2.2. Physiological sensors

We used three wearable devices to collect physiological signals: (1) Pupil Core eye-tracking glasses from Pupil Labs, (2) an H10 chest band from Polar, and (3) an E4 wristband from Empatica. The Pupil Core glasses were used to measure the pupil diameter, dispersion of eye movements, and fixation, sampled at 200 Hz. Then, the H10 chest band was used to collect the heart rate (HR) and Electrocardiogram at sampling rates of 1 and 130 Hz, respectively. Finally, the E4 wristband was used to sample the galvanic skin response (GSR) at 4 Hz, HR at 1 Hz, and photoplethysmography signals at 64 Hz.

4.3. OEDR task and critical events scenarios

Fig. 2 illustrates the experimental scenario. In our simulation software, the ego vehicle automatically drove at 50 km/h in the right lane of a two-lane one-way traffic road with each lane having a width of 4 m. A leading vehicle, traveling at the same speed as the ego vehicle, was located 27 m ahead. On each sidewalk along the road, pedestrians appear at intervals of 37.6 m. In takeover situations, one of three critical events can occur: (1) sudden braking of the leading vehicle and (2–3) pedestrians jaywalking from either sidewalk. These events started at a random time between 60 and 120 seconds after the ego vehicle started. Since events can occur from either the front or both sides of the vehicle, drivers had to monitor the entire road, including the front and both sides of their vehicle. Each critical event lasted approximately two seconds to complete (i.e., collision with the ego vehicle). For example, after the leading vehicle stops, it takes two seconds for the ego vehicle to crash into it. When a critical event occurred, the drivers need to take back control of the vehicle (i.e., takeover) to avoid collisions.

4.4. Auditory secondary tasks

The participants performed three naturalistic tasks and three standardized (or pseudo) tasks in random order during the simulated L2

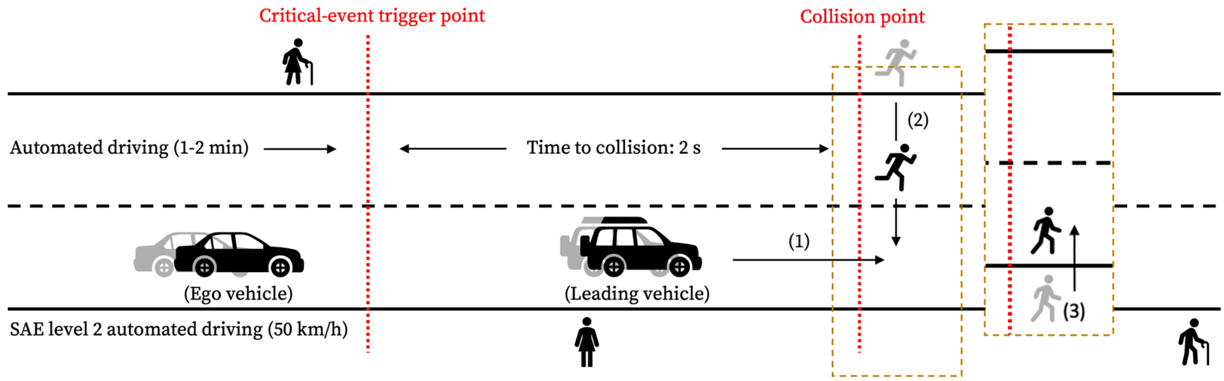

Fig. 2. Three critical-event scenarios: (1) sudden braking of the leading vehicle and jaywalking of pedestrians from the (2) left sidewalk and (3) right sidewalk.

automation. They used a speakerphone mounted behind the steering wheel to perform the secondary tasks (see Fig. 1). Once a secondary task was completed, the next task was started immediately to keep the participants engaged in secondary tasks.

4.4.1. Naturalistic tasks

For naturalistic tasks, we considered both uni- and bi-directional interactions, considering that task demand levels can vary according to interaction types (Lee et al., 2017, Rann and Almor, 2022, Horrey et al., 2009). For the uni-directional interaction task, we selected audiobook listening. For the bi-directional interaction task, we considered auditory texting and auditory gaming.

- *Audiobook listening*: A voice assistant verbally presented a 3-minute excerpt from a farce story called "A Marriage Proposal" (Chekov, 1890). Before the task, the drivers listened to a different excerpt from the same story to get accustomed to the speed and tone of the reader.
- *auditory texting*: The researcher asked the drivers about their daily lives (e.g., "What did you eat last night?"). Then, the drivers listened to a text message and replied to the assistant in the form of voice commands. When the researcher sends a question, a notification sounds from the speaker. The driver then listens to the question and responds to the researcher using the voice assistant. The voice commands were similar to those of Google Assistant with Android Auto. This task required the linguistic resources of the driver.
- *auditory gaming*: The drivers played a classic guessing game called "Guess the Number" (Binder et al., 2021), in which they attempted to guess a randomly selected number between 1 and 100 in a minimum number of attempts. The voice assistants gave feedback by saying "Up" or "Down" if the answer was larger or smaller than the number guessed by the driver, respectively; if the answer was correct, the assistant said, "Correct." This task requires the mental arithmetic resources of the driver.

4.4.2. Standardized tasks: n-back tasks

While naturalistic tasks are realistic and likely to be performed in a daily driving scenario, they lack comparability and repeatability because there is no clear standard for such tasks. For example, while we employed *Guess the Number* for auditory gaming, different types of speech-based games may be employed in other studies. Therefore, we additionally considered *n*-back (delayed digit recall) tasks as standardized tasks (Mehler et al., 2011). Typically, 0- (low demand), 1- (moderate demand), and 2-back (high demand) tasks were utilized.

For the *n*-back task, we followed the standard procedure described in (Mehler et al., 2011). Our conversational agents sequentially presented ten randomly selected numbers from 0 to 9 at 2.25-second intervals. When each number was presented, the drivers were asked to respond verbally with the last n^{th} digit. For example, in the 0-back task, drivers repeated each number after it was presented. In the 2-back task, drivers repeated the third-to-last number in the sequence.

In addition to having high comparability and repeatability, these *n*-back tasks allowed us to objectively measure the task performance (or accuracy), which is defined as the ratio of the number of correctly answered items to the total number of items. For example, if a person correctly answered seven items for the 0-back task, the accuracy was 0.70.

4.4.3. Measuring demands of auditory secondary tasks

Before conducting our main study, we conducted a separate study to analyze how our auditory tasks varied according to the task demands. For this study, we recruited 25 participants (18 males and 7 females) with a mean age of 22.7 years (SD = 1.7 years). They were asked to perform each auditory task for two minutes. Once each task was completed, they assessed the task demand using NASA-TLX (Hart and Staveland, 1988) by following a procedure that includes a weighting stage, as described in (Hart and Staveland, 1988). Participants were each compensated approximately 10 USD.

The average task demands increased in the following order: 0-back (M=10.9, SD=12.9), audiobook listening (M=11.0, SD=11.0), auditory gaming (M=20.1, SD=14.5), auditory texting (M=28.6, SD=17.2), 1-back (M=34.6, SD=23.0), and 2-back (M=57.3, SD=21.9). As shown in Fig. 3, we also statistically

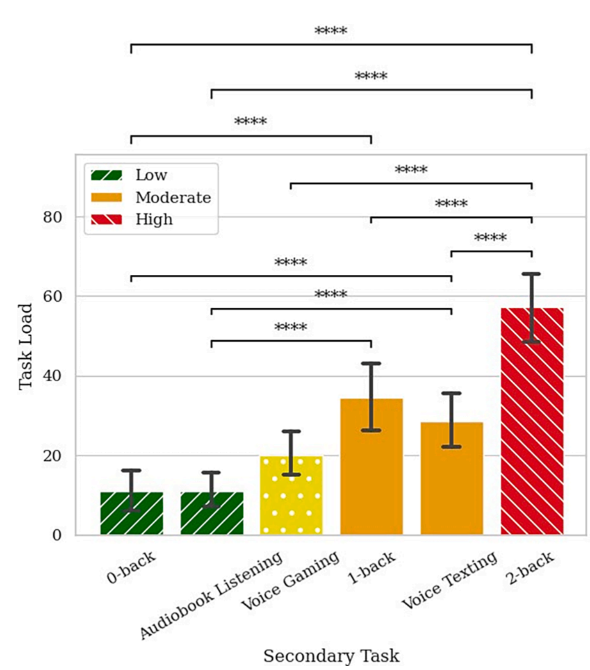

Fig. 3. Distribution of secondary tasks' demands and their pair-wise comparisons. (*p < .05, **p < .01, ***p < .001, ****p < .0001).

analyzed whether task demands differed by task type using a Repeated Measures ANOVA and found significant differences in task demands ($F(3.32, 79.75) = 40.78$, $p < .001$, $\eta^2 = 0.47$). In addition, our post-hoc analysis revealed three distinct groups for which task demands belonging to one group were statistically different from those of the other groups: *low-level* (0-back, audiobook listening), *moderate-level* (1-back, auditory texting), and *high-level* (2-back). auditory gaming was not statistically differentiated between low and moderate levels. Therefore, we did not include auditory gaming in any level groups and further excluded it from RQ1 analyses. For RQ2 analyses, we considered all six auditory secondary tasks since we needed to explore the relationship between physiological context and the takeover performance for every secondary task, irrespective of the difficulty group of the secondary task.

4.5. Measurements

4.5.1. Takeover performance measurements

We considered both time-based and takeover quality measures. These measures were calculated using simulation logs (see Section 4.2.1). For the time-based measure, we considered the takeover reaction time. The reaction time is defined as the elapsed time from the onset of a critical event (e.g., jaywalking) to the first moment of the reaction of a driver (e.g., braking or steering to take control of the vehicle). While there were no instances of false reactions, in two cases, drivers failed to respond to takeover situations while engaging in auditory secondary tasks. In these cases, the reaction time was set as the time between the occurrence of a critical event and the first moment of the crash (i.e., 2 seconds).

Time-based measures have limitations in indicating the safety levels in a takeover situation, because other factors (e.g., type of takeover strategy and the complexity of the road) can also significantly impact the safety levels in such situations (Louw et al., 2017, Blommer et al., 2017). Therefore, to gain a more comprehensive understanding of the safety levels during takeovers, we incorporated a takeover quality measure, specifically the takeover success. The takeover success refers to whether a driver successfully takes the control of a vehicle and avoids a collision in a critical event. It has a binary outcome: success or failure. Takeover was set as a failure if the ego vehicle crashed into other objects (e.g., the leading car or pedestrian) or drove onto a sidewalk.

4.5.2. Physiological measurements

We considered physiological features commonly appearing in previous works involving multitasking and driver distraction (Du et al., 2020, Mehler et al., 2009, Henelius et al., 2009, Kun et al., 2013, Solhjoo et al., 2019). In detail, we obtained the galvanic skin response (GSR), heart rate (HR), inter-beat interval (IBI), pupil diameters, dispersion of eye movements, and off-road glance rate from wearable sensor data (see Section 4.2.2). We then calculated the physiological contexts from measurements collected within a 10-second window preceding a critical event. For example, the off-road glance rate was derived by dividing the number of times there were visual fixations away from the center screen of the cockpit module by the number of times there were whole fixations. The other measurements within a 10 second window were aggregated to calculate the mean and SD. We note that the choice of a 10-second window duration was based on an empirical observation that gave the most significant result for our analyses among varying window durations used in prior studies (Solovey et al., 2014, Liang et al., 2007).

4.6. Recruitment and participants

We recruited 50 drivers who had valid driving licenses. Drivers who had not driven in the last month or did not drive at least once a week were excluded from the recruitment pool. Additionally, for the smooth progression of the experiment, participants who were not familiar with voice assistants were also excluded from the pool. Given that the criteria of the U.S. National Highway and Transportation Safety Administration (NHTSA) for distracted driving studies (NHTSA, 2012), we considered a wide age range and equal numbers of males and females to improve the generalizability. For each age group (i.e., twenties, thirties, forties, fifties, and sixties), there were ten drivers (5 females and 5 males). The average age was 43.8 years (SD = 14.3 years, range = 20–67 years), and the average amount of driving experience was 16.9 years (SD = 11.8 years, range = 1–43 years). The drivers were compensated approximately 30 USD for participation.

4.7. Procedures

The experiment lasted for 120–180 minutes with a 40-minute break. After signing the Institutional Review Board (IRB) consent form, the drivers completed a brief questionnaire asking about their demographic information (e.g., age, driving experience) and their experience with voice assistance and auditory tasks (e.g., audiobook listening or speech-based texting). The drivers then wore three wearable devices (eye-tracking glasses, a H10 chest band, and an E4 wristband; for details, see Section 4.2.2). Next, we briefed the drivers about the simulator and takeover situation and asked them to practice a series of driving trials to familiarize themselves with the simulated automated driving and takeover situations. They performed at least four driving trials and were allowed to perform additional trials. On average, the driving trials lasted for five minutes each.

Next, the drivers experimented with seven different scenarios in which they performed an OEDR task (i.e., monitoring driving environments and then manually taking over control of the vehicle to avoid a collision) while engaging in auditory secondary tasks during L2 automated driving. Each scenario lasted for 1.5 minutes on average (SD = 21.6, range = 1–2). The orders of the scenarios and auditory secondary tasks were counterbalanced by balanced Latin-square design. Before each scenario, we explained a corresponding secondary task and asked the drivers to continue performing the task trials until they were familiar with the tasks. The task trials lasted for an average of 7 minutes; this time varied depending on how quickly drivers familiarized themselves with the tasks. After each scenario, the drivers were given a 5-minute break.

4.8. Analysis method

We conducted statistical analyses for our two research questions. For RQ1, we conducted two mixed-model analyses. In each analysis, we statistically analyzed how the takeover performance differed when the driver simultaneously performed an OEDR task and an auditory secondary task (e.g., listening to an audiobook) and when they only performed the OEDR task (i.e., baseline condition). As shown in Table 3, for the dependent variables, we used takeover reaction time and takeover success. The fixed effects among independent variables were the types of auditory secondary tasks. Additionally, we included drivers and critical-event types (crossed with drivers) as random effects to control the non-independence of the data. A binary logistic regression model was employed to analyze the takeover success because this variable had binary (dichotomous) outcomes. The reaction time, which was in the form of continuous numeric outcomes, was modeled by linear regression. For additional analyses, we transcribed the interview records and used affinity diagramming.

Table 3
Description of dependent variables.

Dependent variables	Descriptions
Reaction time	The elapsed time from the onset of a critical event to the first moment of the reaction of a driver.
Takeover success	Whether a driver successfully takes control of a vehicle and avoids a collision in a critical event.

Table 4
Description of independent variables for Hypothesis 2.

Independent variables	Descriptions
Pupil diameter M	Mean for 10 seconds of the average pupil diameter of the driver's eyes.
Pupil diameter SD	Standard deviation for 10 seconds of the average pupil diameter of the driver's eyes.
Dispersion M	Mean for 10 seconds of the gazes dispersion within a fixation.
Dispersion SD	Standard deviation for 10 seconds of the gaze dispersion gazes within a fixation.
Off-road glance rate	The division between the number of times that fixations were away from the center screen of the cockpit module by the number of whole fixations.
IBI M	Mean for 10 seconds of the time interval between successive heartbeats.
IBI SD	Standard deviation for 10 seconds of the time interval between successive heartbeats.
GSR M	Mean for 10 seconds of the electrical conductance of the skin, which varies with its moisture level.
GSR SD	Standard deviation for 10 seconds of the electrical conductance of the skin, which varies with its moisture level.

For RQ2, we also conducted two multilevel regression analyses to minimize the effects of outliers and erroneous assumptions regarding the shape of the distribution (Field and Wilcox, 2017). In each analysis, we statistically analyzed how each physiological contextual factor accounted for the reaction time or takeover success. For the dependent variables, we used takeover reaction time and takeover success. The independent variables were the physiological contexts. Table 4 shows the descriptions of the physiological contexts used as the independent variable. Before the analyses, we checked whether the independent variables satisfied the primary assumptions of the analyses (e.g., normality, homoscedasticity, and multicollinearity). In this process, we excluded features such as the mean and SD of the HR because they had violated the assumption of multicollinearity (Thompson et al., 2017).

5. Results

5.1. RQ1: how varying levels of auditory secondary tasks affect the takeover performance?

In this section, we reported the descriptive statistics of the experimental results and statistically analyzed how takeover performance varies with different types of auditory–verbal secondary tasks. Additionally, based on the results of the statistical analysis and interviews, we conducted a follow-up analysis to statistically examine drivers' auditory task performance.

5.1.1. Descriptive statistics

The repeated-measures correlation (Bakdash and Marusich, 2017) between the reaction time and the takeover success was negatively moderate ($r = -0.43$, $p < .001$), indicating that a longer reaction time corresponded to a lower takeover success. Considering the average takeover performance, the baseline (no secondary task) condition had the shortest reaction time and the highest takeover success (see Table 5). Except for the baseline condition, the reaction time was the shortest for 0-back and the longest for auditory texting. Conversely, the takeover success was the highest for 0-back and the lowest for auditory texting and 1-back.

Table 5
Descriptive statistics of takeover performance.

	Reaction time (s)		Takeover success	
Secondary task	Mean	SD	Percentage	n
Baseline	0.88	0.20	86	43
Low level				
0-back	0.88	0.24	82	41
Audiobook listening	0.91	0.21	76	38
Moderate level				
1-back	0.96	0.25	68	34
Auditory texting	1.02	0.26	68	34
High level				
2-back	0.98	0.21	72	36

5.1.2. Takeover performance

We statistically analyzed how the reaction time and takeover success varied depending on the task conditions compared to the baseline condition. Our results indicated that auditory secondary tasks degraded the takeover performance and that their effect on the performance varied according to the task demands. As shown on the left side of Table 6, the takeover reaction time was significantly longer for the moderate- (1-back: $\beta = 0.08$, $p = .014$; auditory texting: $\beta = 0.11$, $p < .001$) and high-level (2-back: $\beta = 0.11$, $p < .001$) conditions than for the baseline. In addition, the post-hoc analysis indicated that the reaction times did not differ significantly among the moderate- and high-level conditions (1-back vs. 2-back: mean diff. = 0.034, $t = 1.086$, $p = .279$; auditory texting vs. 2-back: mean diff. = 0.004, $t = 0.134$, $p = .894$). Next, as shown on the right side of Table 6, th1e takeover success was only significantly lower for the moderate-level condition (1-back: $\beta = -1.14$, $OR = 0.32$, $p = .036$; auditory texting: $\beta = -1.14$, $OR = 0.32$, $p = .036$) than for the baseline condition.

Our analysis revealed different results from the H1.a. There was no statistically significant difference in takeover success between the high-level and baseline conditions. Interestingly, the moderate-level condition showed a statistically significant difference compared to the baseline condition. To understand these results, after the analysis, we conducted an additional interview with the following questions: (1) How did you engage in the scenario when L2 driving without secondary tasks versus with secondary tasks? (2) How did you engage in the scenario during L2 driving depending on the different types of secondary tasks? Seventeen drivers agreed to the interview and provided insightful responses. Our interviews revealed that 76% of the drivers we interviewed (n=13/17) intentionally paid less attention to the high-level task and more attention to the OEDR task. For example, P27 stated, "When I was doing the 2-back task, I focused more on the road because I was concerned about avoiding people jumping out onto the road." Some drivers indirectly mentioned a performance trade-off between the primary and secondary tasks. For example, P18 stated, "I felt 2-back was far more difficult than 1-back. (…) When I was doing the 2-back, I focused more on checking the road. I thought it would cause an accident if I didn't do so because 2-back was too difficult. (…) I got an accident when I was doing 1-back because I thought 1-back was not difficult, so I engaged in both 1-back and checking an event."

According to our interview analysis, one possible explanation for our results could be the performance trade-off between the primary task (i.e., the OEDR task) and the secondary task (i.e., the 2-back task). Fuller's Task-difficulty homeostasis model suggests drivers maintain a consistent level of task difficulty and adjust their behavior to balance the level (e.g., reducing the difficulty when the level becomes high) (Fuller, 2005). Similarly, in our study, as soon as facing difficulty to simultaneously perform the OEDR and 2-back tasks, the drivers may adjust their behavior, resulting in the performance trade-off. In other words, drivers may have reduced their 2-back task performance to avoid collisions.

5.1.3. Follow-up analyses

We conducted follow-up analyses to validate our explanation of

Table 6
Mixed-model analysis results for the effects of different secondary tasks on the takeover performance (*$p < .05$, **$p < .01$, ***$p < .001$).

	Reaction time					Takeover success				
Predictors	β (SE)	t-value	95% CI	p		β (SE)	z-value	Odds ratio (95% CI)	p	
(Intercept)	0.87 (0.02)	35.00	0.82 ~ 0.92	< .001	***	1.99 (0.46)	4.36	7.31 (2.98 ~ 17.95)	< .001	***
Low-level										
0-back	0.02 (0.03)	0.50	-0.04 ~ 0.07	.615		-0.31 (0.58)	-0.54	0.73 (0.23 ~ 2.31)	.593	
Audiobook listening	0.04 (0.03)	1.48	-0.01 ~ 0.10	.139		-0.73 (0.56)	-1.32	0.48 (0.16 ~ 1.44)	.188	
Moderate-level										
1-back	0.08 (0.03)	2.46	0.02 ~ 0.14	.014	*	-1.14 (0.54)	-2.11	0.32 (0.11 ~ 0.93)	.036	*
Auditory texting	0.11 (0.03)	3.42	0.05 ~ 0.17	< .001	***	-1.14 (0.54)	-2.11	0.32 (0.11 ~ 0.93)	.036	*
High-level										
2-back	0.11 (0.03)	3.56	0.05 ~ 0.17	< .001	***	-0.93 (0.55)	-1.68	0.40 (0.13 ~ 1.17)	.094	

driver behavior changes in the high-level condition. In this section, we analyzed two aspects of behavior changes: (1) secondary task behavior and (2) takeover behavior.

Behavior Changes in a Secondary Task: We first analyzed driver behavior changes in secondary tasks. If drivers intentionally paid less attention to secondary tasks in the high-level condition, the performance of the secondary tasks would be lower for the *multitasking condition* (i.e., performing multitasking of the OEDR and secondary tasks) than for the *single-task condition* (i.e., performing only the secondary task). In contrast, in moderate and low-level conditions, there would be no significant difference between the *multitasking condition* and the *single-task condition.*

In our experiment, for each scenario, the drivers first performed secondary task trials to familiarize themselves with the tasks (see Section 4.7). We statistically compared how the *n*-back task performance differed under multitasking and single-task conditions across *n*-back types. The *n*-back tasks induced systematically structured levels of task demand in drivers according to the number of delayed digits, allowing us to systematically compare the task performance across task demand levels. We assumed that *n*-back task performances would be different if changes in secondary-task behaviors occurred.

For the comparison, we conducted a repeated measures ANOVA. The task performance (average accuracy) was included as a dependent variable, and the *n*-back type (1-back vs. 2-back) and task condition (*single-task condition* vs. *multitasking condition*) were included as independent variables. To evaluate the task performance, we considered the average *n*-back accuracy (the number of correct answers divided by the number of required answers) for the last two attempts in the single-task condition and for the last two attempts before the occurrence of a critical event in the multitasking condition. Only the 1-back and 2-back tasks were used for the *n*-back type. The 0-back task was excluded because 0-back was the same as 100% accuracy in both task conditions.

Fig. 4 shows the *n*-back task performances under the *single-task condition* and the *multitasking condition* for the *n*-back type. As shown in Table 7, our analysis indicated that both the main effect ($p < .001$) and the interaction effect were significant ($p = .042$). Furthermore, the post-hoc analysis indicated that the accuracy for the 2-back task was significantly lower in the main task condition than in the trial task condition (mean diff. $= 0.103, t = 3.81, p < .001$). In contrast, the 1-back accuracy exhibited no significant difference (mean diff. $= 0.036, t = 0.18, p = .71$) between the main and trial task conditions. This result supported our explanation of driver behavior changes in the 2-back condition since the difference in the secondary-task performance between the main and trial task conditions was only significant for the 2-back condition.

Behavior Changes in a Takeover Task: Next, we analyzed driver behavior changes in takeover tasks. If drivers intentionally paid less attention to secondary tasks and more attention to the OEDR task in the high-level condition, takeover task behaviors would become more efficient as the performance of secondary tasks decreases. In contrast, in other conditions, takeover task behaviors would become more efficient as the performance of secondary tasks increases, because drivers paid their attention to both tasks. As shown in Fig. 5, our drivers exhibited

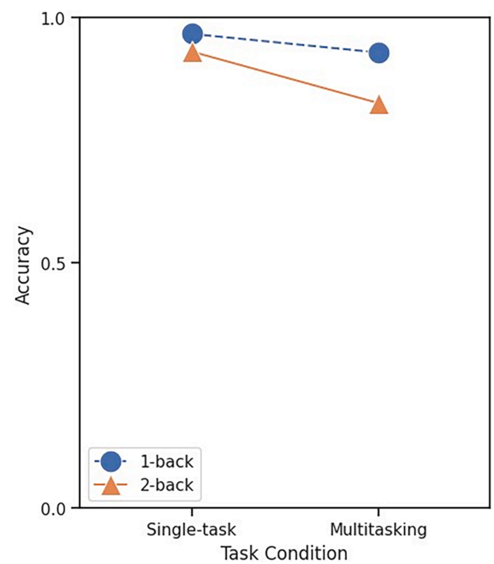

Fig. 4. Accuracy of 1-back and 2-back task.

Table 7
Repeated measures ANOVA results for the n-back accuracy

Effects	F	p	η^2
Task condition	15.96	.001	0.25
n-back type	18.09	.001	0.27
Task condition x *n*-back type	4.37	.042	0.08

two types of takeover tasks: (1) braking with steering (*combined operating*) and (2) braking only (*single operating*). Given that *combined operating* leads to more successful collision avoidance compared to *single operating* (Louw et al., 2017, Rice and Dell'Amico, 1974, Lechner and Malaterre, 1991), for the high-level condition, we assumed that drivers would be more likely to exhibit *combined operating* for takeover tasks as the performance of secondary tasks decreases.

For the analysis, we conducted a binary logistic mixed-model analysis to examine how *n*-back accuracy influenced the probability of drivers engaging in *combined operating* across different *n*-back types. We assumed that *combined operating* would occur more frequently in the high-level condition as *n*-back accuracy decreased. Similar to our secondary task behavior analysis, we excluded the 0-back task from the analysis since its accuracy did not change and included only the 1-back and 2-back tasks for the *n*-back type. The dependent variable was the takeover task. The independent variables were the *n*-back types, *n*-back accuracy, and the interaction between them. We also included drivers and critical-event types as random effects to control for the non-independence of the data.

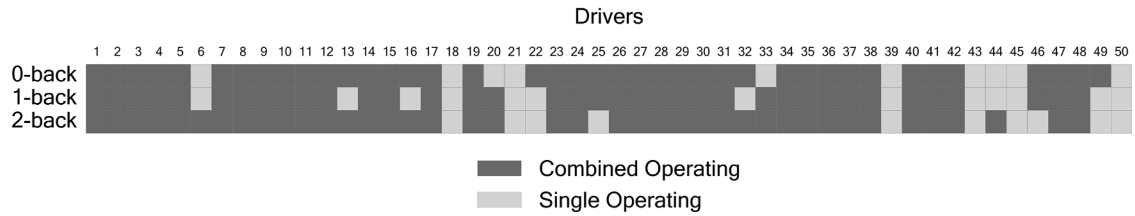

Fig. 5. A heatmap of drivers' takeover task behavior across *n*-back type.

Table 8
Mixed-model analysis results for the effects of *n*-back type and *n*-back accuracy on the takeover task (*$p < .05$, **$p < .01$, ***$p < .001$, ****$p < .0001$).

	Takeover behavior (*combined operating*)						
Predictors	β	(SE)	*z*-value	Odds ratio (95% CI)		*p*	
(Intercept)	60.6	-0.002	28,943	6.98×10^{26}	$(4.44 \times 10^{26} \sim 1.10 \times 10^{27})$	$< .0001$	****
Main effect							
n-back type (1-back)	-56.5	-0.002	-26,977	2.27×10^{-25}	$(1.44 \times 10^{-25} \sim 3.57 \times 10^{-25})$	$< .0001$	****
n-back accuracy	-39.3	-0.002	-18,757	8.64×10^{-18}	$(5.49 \times 10^{-18} \sim 1.36 \times 10^{-17})$	$< .0001$	****
Interaction effect							
1-back × *n*-back accuracy	46.5	-0.002	22,174	1.12×10^{20}	$(7.15 \times 10^{19} \sim 1.74 \times 10^{20})$	$< .0001$	****

As shown in Table 8, all independent variables were statistically significant. The significance of the *n*-back types (1-back) and the negative beta coefficient ($\beta = -56.5$, $p < .0001$) suggest that drivers were more likely to exhibit *combined operating* in the 2-back condition than in the 1-back condition. The significance of the *n*-back accuracy and the negative beta coefficient ($\beta = -39.3$, $p < .0001$) suggest that drivers were more likely to exhibit *combined operating* as *n*-back accuracy decreased. Additionally, the significance of the interaction effect and positive beta coefficient ($\beta = 46.5$, $p < .0001$) indicate that in the 1-back condition, drivers were more likely to exhibit *combined operating* as *n*-back accuracy increased.

While the results of our follow-up analyses supported our explanation, further study in-depth research is required to rigorously investigate into the impacts of behavior changes on takeover performance. In Section 6.1 and Section 6.5, we further discussed these findings and future research directions.

5.2. RQ2: which are the typical physiological contexts related to the takeover performance?

In this section, we statistically analyze which physiological measures are related to takeover performance.

5.2.1. Physiological contexts

Our analyses indicated that the pupil diameter, eye movement dispersion, and IBI had statistically significant correlations with the takeover performance. Specifically, as shown on the left side of Table 9, pupil diameter SD, and eye-movements dispersion SD were significant. Given that the coefficient value of the pupil diameter SD was positive ($\beta = 0.06$, $p = .025$), drivers were more likely to exhibit longer reaction times when they had higher variability in the pupil diameter before performing a takeover task. Considering the negative coefficient value of the eye-movements dispersion SD ($\beta = -0.19$, $p = .014$), the reaction times were more likely to be shorter when the drivers had higher variability in the dispersion of eye movements. As shown on the right side of Table 9, the mean ($\beta = 5.62$, $OR = 276.95$, $p = .010$) and SD ($\beta = 6.21$, $OR = 496.37$, $p = .017$) of the eye-movements dispersion and IBIs SD were significant. The coefficient values of the mean and SD of the eye-movements dispersion were positive, indicating that drivers were more likely to succeed when the dispersion of eye movements was wider or more varied. However, the coefficient values of the IBIs SD were negative ($\beta = -1.61$, $OR = 0.20$, $p = .012$), indicating that drivers were more likely to succeed in takeovers when the IBI SD was lower. The GSR M, GSR SD, and off-road glance rate were not significant for the takeover performance.

6. Discussion

Auditory interfaces have been widely deployed in-vehicle systems with different levels of automation, from non-automated to conditionally automated vehicles (L0–L3). We conducted a quantitative study to investigate the impact of in-vehicle auditory interactions on takeover performance in L2 automated driving contexts. The results showed that in-vehicle auditory interactions could negatively impact the takeover performance, and that this impact varied depending on the task-demand level. We also found that the takeover performance was significantly associated with physiological contexts such as pupil diameter, dispersion of eye movements, and inter-beat interval (IBI). In the following, we

Table 9
Mixed-model analysis results for the effects of different physiological measures on the takeover performance ($p < .05$, **$p < .01$, ***$p < .001$).

	Reaction time					Takeover success				
Predictors	β (SE)	*t*-value	95% CI	*p*		β (SE)	*z*-value	Odds ratio (95% CI)	*p*	
(Intercept)	0.92 (0.01)	79.82	0.90 ~ 0.95	< .001	***	1.54 (0.20)	7.57	4.67 (3.13 ~ 6.96)	< .001	***
Pupil diameters M	-0.08 (0.04)	-1.78	-0.17 ~ 0.01	.076		-0.11 (0.71)	-0.16	0.90 (0.22 ~ 3.64)	.877	
Pupil diameters SD	0.06 (0.02)	2.25	0.01 ~ 0.10	.025	*	-0.27 (0.46)	-0.58	0.77 (0.31 ~ 1.90)	.565	
Dispersion M	-0.11 (0.07)	-1.60	-0.24 ~ 0.03	.112		5.62 (2.16)	2.61	276.95 (3.99 ~ 19238.71)	.010	**
Dispersion SD	-0.19 (0.08)	-2.48	-0.34 ~ -0.04	.014	*	6.21 (2.58)	2.41	496.37 (3.12 ~ 79007.83)	.017	*
Off-road glance rate	0.01 (0.01)	1.01	-0.01 ~ 0.03	.316		-0.03 (0.13)	-0.23	0.97 (0.75 ~ 1.25)	.820	
IBI M	-0.01 (0.02)	-0.46	-0.05 ~ 0.03	.650		0.36 (0.36)	0.99	1.43 (0.71 ~ 2.88)	.321	
IBI SD	0.03 (0.02)	1.52	-0.01 ~ 0.08	.131		-1.61 (0.64)	-2.52	0.20 (0.06 ~ 7.03)	.012	*
GSR M	0.03 (0.06)	0.56	-0.08 ~ 0.14	.577		0.66 (0.83)	0.80	1.93 (0.38 ~ 9.82)	.426	
GSR SD	0.00 (0.05)	0.08	-0.09 ~ 0.10	.935		-1.12 (0.69)	-1.63	0.33 (0.09 ~ 1.26)	.105	

further discussed the interpretation of our findings, practical considerations for deploying in-situ physiological measures, and design implications for mitigating the negative impacts of auditory secondary tasks on takeover performance.

6.1. Effects of auditory interactions on takeover performances

Our work highlights that researchers and practitioners should carefully design in-vehicle auditory interactions (or interfaces) for safety in L2 driving contexts. Specifically, our results showed that auditory secondary tasks in takeover situations could significantly increase the time required to maneuver vehicles (i.e., increase in the reaction time) and/or the probability of having a collision (i.e., decrease in the takeover success). Our results contrasted with findings in prior studies involving auditory interactions in L3 automated vehicles, wherein such interactions did not impact the takeover performance (Pakdamanian et al., 2021, Wandtner et al., 2018, Berghöfer et al., 2018). For example, Wandtner et al. (2018) investigated how different modality types of secondary tasks (i.e., auditory–vocal vs. visual–vocal vs. visual–manual) impacted the takeover performance in L3 contexts and found that when compared to the no secondary task condition, the auditory-vocal secondary task did not result in a significant increase in the reaction time in response to a takeover request.

In contrast with results in L3 contexts, our findings were similar to those in prior studies involving auditory interactions in manual driving contexts (Kim et al., 2020, Strayer et al., 2015). For example, Strayer et al. (2015) investigated how the varying demands of auditory secondary tasks impacted the brake reaction time in response to the unpredictable braking of the leading vehicle in manual driving contexts and found that when compared to the condition without a secondary task, the auditory task with a high demand resulted in a greater increase in the brake reaction time. Specifically, similar to our results, Strayer et al. found no difference in the brake reaction time when drivers had been engaged in uni-directional driver-vehicle interactions (e.g., audiobook listening). In contrast, brake reaction time was longer when drivers engaged in bi-directional driver-vehicle interactions (or computer-mediated interactions; e.g., auditory texting).

One possible explanation for the similarities and differences in our results compared to previous studies on manual driving (L2 vs. manual driving) and L3 driving (L2 vs. L3 driving) could be the increased workload experienced by drivers when performing an auditory task. This increased workload may result from the need to perform an additional task (i.e., driving tasks in manual driving; OEDR tasks in L2 driving), in addition to the auditory task. In L3 driving contexts, drivers can solely devote their attention on the auditory task until a critical event occurs, without engaging in a primary task (e.g., monitoring task). In contrast, in manual and L2 contexts, drivers must simultaneously allocate attention to both the auditory task and the primary task. Similar to our explanation, the literature also suggests that the overall workload experienced by the drivers when performing secondary tasks is significantly higher for manual and L2 driving than for L3 driving (Figalová et al., 2024).

6.2. Task-demand levels and driver behavior changes

As discussed in Section 2.2, prior studies have not considered the impacts of varying levels of auditory interaction on takeover performance in L2 automated driving contexts. Our results showed that the impacts of auditory secondary tasks on takeover performance varied according to their task-demand level. We initially hypothesized that the higher cognitive demand required by secondary tasks would be associated with a lower takeover performance. Our findings showed that the reaction time for critical events adhered to such expectations, given that the reaction time was significantly higher in the moderate- and high-level conditions than in the baseline condition. However, surprisingly, the takeover success was only significantly reduced in the moderate-level condition. The takeover success was insignificantly different for the high-level and baseline conditions.

Prior studies also have reported similar findings regarding the mismatch between reaction time and takeover success (Louw et al., 2017, Zeeb et al., 2016). For example, Louw et al. (2017) found that imposing additional demands on drivers through secondary tasks (e.g., *n*-back task) or driving environment factors (e.g., heavy/light fog) increased reaction time but did not necessarily affect when drivers initiated a collision avoidance maneuver or the quality of their subsequent vehicle control after the onset of the takeover. In addition, the literature suggests that beyond the reaction time, other factors, such as takeover tactics (e.g., utilizing both the braking and steering wheel), can also influence the success of collision avoidance (Louw et al., 2017, Blommer et al., 2017, Rice and Dell'Amico, 1974, Lechner and Malaterre, 1991).

In the high-level condition, our drivers exhibited behavior changes by intentionally allocating less attention to the secondary task. This behavioral change can be explained by the task-difficulty homeostasis model (Fuller, 2005). This model suggests that drivers maintain a consistent level of task difficulty and adjust their behavior for such maintenance. In our study, the sharp increase in overall task difficulty due to the 2-back task may prompt the drivers' behavior changes. This behavior has also been widely observed in the context of manual driving (Regan et al., 2008). For example, studies involving manual driving have shown that drivers tend to regulate their engagement with secondary tasks to maintain their driving ability (or driving performance) (Kim et al., 2020, Ismaeel et al., 2020). Literature has shown that in manual driving contexts, drivers adjust their secondary task behavior (e.g., reducing attention or ceasing to engage in secondary tasks (Kim et al., 2020, Ismaeel et al., 2020)) and/or change their driving behavior (e.g., reducing speed (Kim et al., 2020, Oviedo-Trespalacios et al., 2017), maintaining a longer headway (Metz et al., 2015), and reducing lane changes (Kim et al., 2020, Fitch et al., 2014)).

In our study, beyond the decrease of secondary-task performance in the high-level condition, we also observed different takeover task behaviors linked to the changes in secondary-task performance, between the high- and moderate-level conditions. Specifically, in the high-level condition, drivers were more likely to exhibit combined operating (i.e., using both the steering wheel and pedals) as a secondary task the performance decreased. Whereas, in the moderate-level condition, the combined operating was more likely to occur as the performance increased. Given that takeover success was insignificantly different between the high-level and baseline conditions, in the high-level condition, the decrease of secondary-task performance–indicating reduced attention on the secondary tasks–might have enabled drivers to pay more attention to a monitoring task and led to more effective tactical decision making before the applying combined operating during a takeover. Literature suggests that the combined operating requires more cognitive resources than either braking or steering alone (Oostwoud Wijdenes et al., 2016). Takeover success was only significantly different for moderate-level and baseline conditions. In the moderate-level condition, it is possible that drivers might have insufficient resources for effective tactical decision-making as their attention remained on secondary tasks; the secondary-task performance was not significantly decreased in the moderate-level condition.

While we observed changes in driver behavior in response to secondary task demands, further investigation is required to conduct a more robust statistical analysis. This would help examine how changes in secondary task behavior (e.g., reduced attention) affect takeover performance and the discrepancy between reaction time and takeover success. Unfortunately, our study may have lacked sufficient statistical power to confirm this effect. In Section 6.5, we address these limitations and suggest future research directions.

6.3. Real-time estimation of takeover performance in L2 driving vehicles

Our results indicated that ocular measures, such as dispersion of eye movements and pupil diameter, were strongly correlated with takeover performance. However, among the physiological features related to ocular measures, the off-road glance rate did not have a significant effect in either analysis. Although prior studies have closely linked the off-road glance rate with takeover performance (Louw et al., 2019, Berghöfer et al., 2018), they mainly considered the context of visual–manual tasks, which led drivers to share their visual operations with the monitoring task (i.e., off-road glance behavior). Unlike these studies, our study considered auditory tasks that did not require visual sharing with the monitoring task. Therefore, the off-road glance behavior would be less likely to occur and not vary across task conditions, leading to statistical insignificance in our analyses.

Currently, commercial automated vehicles (e.g., Tesla and Ford) monitor the off-road glance behavior of drivers and restrict the activation of automated driving features if the off-road glance behavior is monitored (Barry, 2022). However, our results highlighted that the current monitoring approach may not effectively monitor the degradation of the takeover performance caused by in-vehicle auditory interactions. Instead of physiological contexts related to off-road glance behavior, other physiological contexts need to be considered to monitor the degradation of the takeover performance in L2 automated driving. Our results showed that pupil diameter, dispersion of eye movements, and IBI were significantly related to the takeover performance of a driver.

Prior studies have suggested various approaches to obtain physiological signals in vehicular environments. For example, Shin et al. (2010) embedded electrocardiogram and photoplethysmography sensors on steering wheels to monitor the physiological conditions of drivers. This approach could effectively collect the physiological responses of drivers in L2 vehicles since the drivers need to constantly have their hands on the wheel to activate the automation mode (Barry, 2022). Alternatively, similar to our study setting, such responses could be collected using wearable devices, such as a smartwatch, by considering the popularity of such devices. Prior studies have shown that the pupil diameter and dispersion of eye movements could be reasonably measured using a low-cost camera (Wangwiwattana et al., 2018, Huynh et al., 2022). A camera could be installed in a vehicle to collect such physiological signals. Indeed, an existing camera in automated vehicles could also be leveraged to collect such signals, given that recent automated vehicles from the major manufacturers (e.g., Tesla and Ford) have started to equip camera-based driver monitoring systems to monitor the off-road glance behavior of the drivers (Barry, 2022).

6.4. Toward context-aware adaptive automated driving

To enable drivers to successfully perform an OEDR task while they engage in auditory interactions, we envision future L2 automated vehicles that automatically adapt their driving style or auditory interactions (i.e., context-aware adaptation approach). The majority of drivers who participated in our interview ($n = 13/17$) reported that they had prioritized safety and exhibited behavior changes in that they had regulated their engagement in secondary tasks to maintain the performance of the OEDR task. However, drivers are not always rational and may suffer from various cognitive biases and errors. For example, they may overestimate their driving/OEDR skill or underestimate the chance to be involved in a driving accident (i.e., optimism bias) (DeJoy, 1989). Furthermore, they may underestimate the potential impact of auditory secondary tasks on their performance in an OEDR task (i.e., risk-taking behaviors) (Fuller, 1991). In our study, drivers did not exhibit behavior changes when they were performing moderate-level auditory tasks, although these tasks degraded their takeover performance. Indeed, overtrust in an imperfect L2 automation system might discourage drivers from executing any behavior changes (Lee and See, 2004, Wagner et al., 2018). Instead of relying on drivers (e.g., behavior changes), the context-aware adaptive automation approach could help assist drivers in such cases.

In L2 automated driving, the system controls both the longitudinal and lateral positions of the vehicle; drivers are not involved in these functions. This non-involvement can be beneficial in normal contexts. However, in our study, some drivers expressed concerns of their non-involvement while engaging in secondary tasks. In manual driving contexts, while performing a secondary task, drivers tend to adjust their driving styles to compensate for the additional workload that is induced by the secondary task. During our interviews, one driver (P14) emphasized this difference, stating that if he was driving, he would "stop the car before performing the task or reduce the speed." He said, "I felt it was not safe (to perform the OEDR task while engaging in auditory secondary tasks) because the (automated) vehicle did not automatically stop or slow down." Future automated driving systems may automatically adjust their driving style while drivers engage in auditory secondary tasks.

Similarly, L2 automated vehicles may automatically reduce their speed gradually, or change lanes to low-speed lanes when the takeover performance of a driver is expected to be low. This deceleration approach can help reduce the overall takeover difficulty for the driver (e.g., slower speed shortens the stopping distance required after braking is initiated), potentially aiding in collision avoidance. In addition to context-aware driving adaptation, we can also consider auditory secondary task adaptation. Kim et al. proposed a secondary task adaptation approach in which in-vehicle agents control auditory interactions (e.g., reducing the length of the auditory interactions) by considering the interruptibility of drivers in manual driving contexts (see Section 6.3 in (Kim et al., 2020)). Such an adaptation approach could also be applied to in-vehicle voice assistants or auditory interfaces for L2 vehicles. For example, in-vehicle voice assistants may decompose an auditory interaction into multiple micro-interactions, pause a micro-interaction, and resume it at opportune moments when the takeover performance is estimated to be high (Kim et al., 2020). Kim et al. (2018) showed that vehicle-context sensor data can be used to find the opportune moments when drivers can safely engage in auditory interactions in manual driving contexts. Similarly, to find the opportune moments, the automated vehicles could monitor the in-situ physiological signals of a driver (e.g., pupil diameter) and estimate their takeover performance, as discussed in Section 6.3.

Context-aware adaptation approaches could be potentially more effective than restrictive approaches. Currently, restrictive approaches are widely applied in commercial automated vehicles; the activation of automated driving features is restricted when drivers are not holding the steering wheel or are inattentive on the road (Barry, 2022). However, such restrictive approaches may not be effective in practice (Creaser et al., 2015, Akuchie, 2023, Klender, 2023). For instance, it was found that some Tesla drivers slept at the wheel while using defeat devices to deceive the driver monitoring system (e.g., monitoring off-road glance behavior) into believing that they were attentive (Akuchie, 2023). In addition, such restrictive approaches may negate the potential benefits of providing auditory secondary tasks to drivers experiencing under-load situations (or low levels of mental workload) (Mishler and Chen, 2024, Vogelpohl et al., 2019). Literature suggests that auditory secondary tasks can help drivers maintain their optimal performance in an OEDR task when they solely perform a monitoring task alone over long periods of time (Mishler and Chen, 2024), or experience drowsiness (Vogelpohl et al., 2019). In such challenging scenarios, our findings and the concept of context-aware adaptive automated driving could help drivers maintain optimal performance in OEDR tasks.

6.5. Limitations and future work

Although our study demonstrated that different task demands could influence takeover performance in L2 automated driving contexts, our results should be carefully interpreted and generalized when applied to real-world L2 automated driving vehicles. First, although we recruited a broad age range and ensured equal gender representation among 50 participants to meet NHTSA guidelines and exceed the sample sizes of previous studies (see Table 2), we acknowledge the need for a larger sample size in future research to ensure sufficient statistical power to detect the effects of secondary tasks on the takeover success. Next, our participants experienced experimental scenarios during a practice session preceding the experiment. Therefore, although they did not know when critical events (e.g., jaywalking) would occur, they could expect that such events might happen at some point and respond accordingly. However, in real-world driving contexts without prior takeover training (or experience), drivers may overestimate their ability to handle sudden incidents due to various cognitive biases (e.g., capability overestimation, risk-taking behaviors, optimism bias) (DeJoy, 1989, Fuller, 1991, BROWN, 1990). As a result, the drivers may not exhibit behavior changes (e.g., not reducing engagement in secondary tasks with high demand). Thus, further research is needed to investigate how driver behavior changes affect takeover performance in various takeover scenarios. Next, further in-depth research is required for a more rigorous investigation into the impact of other factors on takeover performance. The literature has shown that takeover performance can also be influenced by the road environment, such as the complexity of the traffic situation (Radlmayr et al., 2014), and the driver's temporal conditions, such as drowsiness (Naujoks et al., 2018) or emotional state (Sanghavi et al., 2023, Sanghavi et al., 2020, Pan et al., 2024). Future studies should comprehensively consider these factors alongside the secondary task. While we demonstrated the feasibility of estimating takeover performance by analyzing the physiological responses of drivers, further investigation is required to develop reliable predictive models. Unfortunately, in our study, such predictions were not feasible due to an insufficient dataset.

7. Conclusion

We investigated the effects of auditory interaction on the takeover performance during L2 automated driving. Our results showed that the takeover performance was affected by the demands of the auditory interactions, which could be monitored by observing the physiological contexts of the drivers. Based on these findings, we discussed methods for both the real-time estimation of the decreases in takeover performance due to auditory interaction and the mitigation of this decrease. We hope our results can contribute to the safe use of auditory interactions in L2 automated driving contexts.

CRediT authorship contribution statement

Jiwoo Hwang: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation. **Woohyeok Choi:** Writing – original draft, Validation, Supervision, Investigation, Formal analysis. **Auk Kim:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1C1C2003865) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00242528).

Data availability

The authors do not have permission to share data.

References

Akuchie, M., 2023. Tesla's staged self-driving video is a prime example of deceptive marketing. ScreenRant. https://screenrant.com/tesla-staged-self-driving-video-deceptive-marketing/.

Alvarez, I., Alnizami, H., Dunbar, J., Jackson, F., Gilbert, J.E., 2015. Help on the road: Effects of vehicle manual consultation in driving performance across modalities. Int. J. Human-Comput. Stud. 73, 19–29.

Arkonac, S.E., Brumby, D.P., Smith, T., Babu, H.V.R., 2019. In-car distractions and automated driving: A preliminary simulator study. In: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings, AutomotiveUI '19. Association for Computing Machinery, New York, NY, USA, pp. 346–351.

Ayoub, J., Zhou, F., Bao, S., Yang, X.J., 2019. From manual driving to automated driving: A review of 10 years of autoui. In: Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '19. Association for Computing Machinery, New York, NY, USA, pp. 70–90.

Bakdash, J.Z., Marusich, L.R., 2017. Repeated measures correlation. Front. Psychol. 8, 252904.

Banks, V.A., Eriksson, A., O'Donoghue, J., Stanton, N.A., 2018. Is partially automated driving a bad idea? observations from an on-road study. Appl. Ergon. 68, 138–145.

Barry, K., Driver monitoring systems can help you be safer on the road (2022). https://www.consumerreports.org/car-safety/driver-monitoringsystems-ford-gm-earn-points-in-cr-tests-a6530426322/.

Berghöfer, F.L.; Purucker, C.; Naujoks, F.; Wiedemann, K.; Marberger, C., Prediction of take-over time demand in conditionally automated driving-results of a real-world driving study, Proceedings of the Human Factors and Ergonomics Society Europe (2018) 69–81.

Binder, F.J.; Jones, C.R.; Kaufman, R.A.; Lin, N.T. Poole, C.R. Vul, E., Cognitive cost and information gain trade-off in a large-scale number guessing game, in Proceedings of the Annual Meeting of the Cognitive Science Society, Vol. 43, 2021.

Blommer, M., Curry, R., Kochhar, D., Swaminathan, R., Talamonti, W., Tijerina, L., 2015. The effects of a scheduled driver engagement strategy in automated driving. Proc. Human Factors Ergonomics Soc. Annual Meeting 59 (1), 1681–1685.

Blommer, M., Curry, R., Swaminathan, R., Tijerina, L., Talamonti, W., Kochhar, D., 2017. Driver brake vs. steer response to sudden forward collision scenario in manual and automated driving modes. Transp. Res. Part F: Traffic Psychol. Behaviour 45, 93–101.

BROWN, I.D., 1990. Drivers' margins of safety considered as a focus for research on error. Ergonomics 33 (10-11), 1307–1314. https://doi.org/10.1080/00140139008925334.

Bueno, M.; Dogan, E.; Hadj Selem, F.; Monacelli, E.; Boverie, S.; Guillaume, A., How different mental workload levels affect the take-over control after automated driving, in: 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), 2016, pp. 2040–2045.

Chekov, A., A marriage proposal, 1890.

Chen, K.T., Chen, H.Y.W., 2021. Manipulating music to communicate automation reliability in conditionally automated driving: a driving simulator study. Int. J. Human-Computer Stud. 145, 102518.

Council of European Union, Commission implementing regulation (EU) 2022/1426 (aug 2022). https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32022R1426.

Creaser, J.I., Edwards, C.J., Morris, N.L., Donath, M., 2015. Are cellular phone blocking applications effective for novice teen drivers? J. Safety. Res. 54, 75.e29–75.e78 strategic Highway Research Program (SHRP 2) and Special Issue: Fourth International Symposium on Naturalistic Driving Research.

DeJoy, D.M., 1989. The optimism bias and traffic accident risk perception. Accident Anal. Prevent. 21 (4), 333–340.

Dosovitskiy, A.; Ros, G.; Codevilla, F.; Lopez, A.; Koltun, V., CARLA: An open urban driving simulator, in Proceedings of the 1st Annual Conference on Robot Learning, 2017, pp. 1–16.

Du, N., Zhou, F., Pulver, E.M., Tilbury, D.M., Robert, L.P., Pradhan, A.K., Yang, X.J., 2020a. Predicting driver takeover performance in conditionally automated driving. Accident Anal. Prevention 148, 105748.

Du, N., Kim, J., Zhou, F., Pulver, E., Tilbury, D.M., Robert, L.P., Pradhan, A.K., Yang, X.J., 2020b. Evaluating effects of cognitive load, takeover request lead time, and traffic density on drivers' takeover performance in conditionally automated driving. In: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '20. Association for Computing Machinery, New York, NY, USA, pp. 66–73.

Du, N., Zhou, F., Tilbury, D.M., Robert, L.P., Yang, X.Jessie, 2024. Behavioral and physiological responses to takeovers in different scenarios during conditionally automated driving. Transportation Res. Part F: Traffic Psychol. Behaviour 101, 320–331. https://doi.org/10.1016/j.trf.2024.01.008.

Eriksson, A., Petermeijer, S.M., Zimmermann, M., de Winter, J.C.F., Bengler, K.J., Stanton, N.A., 2019. Rolling out the red (and green) carpet: Supporting driver

decision making in automation-to-manual transitions. IEEe Trans. Hum. Mach. Syst. 49 (1), 20–31.
Faure, V., Lobjois, R., Benguigui, N., 2016. The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. Transp. Res. Part F: Traffic Psychol. Behaviour 40, 78–90.
Feldhütter, A., Gold, C., Schneider, S., Bengler, K., et al., 2017. How the duration of automated driving influences take-over performance and gaze behavior. In: Schlick, C.M., Duckwitz, S., Flemisch, F., Frenz, M., Kuz, S., Mertens, A., et al. (Eds.), Advances in Ergonomic Design of Systems, Products and Processes. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 309–318.
Field, A.P., Wilcox, R.R., 2017. Robust statistical methods: a primer for clinical psychology and experimental psychopathology researchers. Behav. Res. Ther. 98, 19–38 best Practice Guidelines for Modern Statistical Methods in Applied Clinical Research.
Figalová, N., Bieg, H.J., Reiser, J.E., Liu, Y.C., Baumann, M., Chuang, L., Pollatos, O., 2024. From driver to supervisor: Comparing cognitive load and eeg-based attentional resource allocation across automation levels. Int. J. Human-Computer Stud. 182, 103169. https://doi.org/10.1016/j.ijhcs.2023.103169.
Fitch, G.M., Grove, K., Hanowski, R.J., Perez, M.A., 2014. Compensatory behavior of drivers when conversing on a cell phone: Investigation with naturalistic driving data. Transp. Res. Rec. 2434 (1), 1–8.
Fuller, R., 1991. Behavior analysis and unsafe driving: Warning—learning trap ahead! J. Appl. Behav. Anal. 24 (1), 73.
Fuller, R., 2005. Towards a general theory of driver behaviour. Accident Anal. Preven. 37 (3), 461–472. https://doi.org/10.1016/j.aap.2004.11.003.
Gold, C., Berisha, I., Bengler, K., 2015. Utilization of drivetime – performing non-driving related tasks while driving highly automated. Proc. Human Factors Ergonomics Society Annual Meeting 59 (1), 1666–1670.
Gold, C., Körber, M., Lechner, D., Bengler, K., 2016. Taking over control from highly automated vehicles in complex traffic situations: the role of traffic density. Hum. Factors 58 (4), 642–652.
Hart, S.G.; Staveland, L.E., Development of nasa-tlx (task load index): Results of empirical and theoretical research, in P. A. Hancock, N. Meshkati (Eds.), Human Mental Workload, Vol. 52 of Advances in Psychology, North-Holland, 1988, pp. 139–183.
Henelius, A.; Hirvonen, K.; Holm, A.; Korpela, J.; Muller, K., Mental workload classification using heart rate metrics, in 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009, pp. 1836–1839.
Horrey, W., Lesch, M., Garabet, A., 2009. Dissociation between driving performance and drivers' subjective estimates of performance and workload in dual-task conditions. J. Safety. Res. 40 (1), 7–12.
Huynh, S., Balan, R.K., Ko, J., 2022. Imon: appearance-based gaze tracking system on mobile devices. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 5 (4) dec.
Ismaeel, R., Hibberd, D., Carsten, O., Jamson, S., 2020. Do drivers self-regulate their engagement in secondary tasks at intersections? an examination based on naturalistic driving data. Accident Anal. Prevent. 137, 105464.
Janssen, C.P., Iqbal, S.T., Kun, A.L., Donker, S.F., 2019. Interrupted by my car? implications of interruption and interleaving research for automated vehicles. Int. J. Human-Computer Stud. 130, 221–233. https://doi.org/10.1016/j.ijhcs.2019.07.004.
Kim, A., Choi, W., Park, J., Kim, K., Lee, U., 2018. Interrupting drivers for interactions: Predicting opportune moments for in-vehicle proactive auditory-verbal tasks. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2 (4) dec.
Kim, A., Park, J.M., Lee, U., 2020. Interruptibility for in-vehicle multitasking: Influence of voice task demands and adaptive behaviors. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 4 (1) mar.
Klender, J., 2023. Tesla with sleeping driver proves there's still misunderstanding and irresponsibility surrounding autonomy. TeslaRati. https://www.teslarati.com/tesla-sleeping-driver-autonomymisunderstanding-video/.
Kun, A., Palinko, O., Medenica, Z., Heeman, P., 2013. On the feasibility of using pupil diameter to estimate cognitive load changes for in-vehicle spoken dialogues. In: Proceedings of the Annual Conference of the International Speech Communication Association. INTERSPEECH, pp. 3766–3770, 14th Annual Conference of the International Speech Communication Association, INTERSPEECH 2013; Conference date: 25-08-2013 Through 29-08-2013.
Lechner, D., Malaterre, G., 1991. Emergency manuever experimentation using a driving simulator. Tech. Rep. SAE Technical Paper.
Lee, J.D., See, K.A., 2004. Trust in automation: Designing for appropriate reliance. Hum. Factors 46 (1), 50–80.
Lee, A.M.C., Cerisano, S., Humphreys, K.R., Watter, S., 2017. Talking is harder than listening: The time course of dual-task costs during naturalistic conversation. Canadian J. Exper. Psychol./Revue canadienne de psychologie expérimentale 71, 111–119.
Liang, Y., Reyes, M.L., Lee, J.D., 2007. Real-time detection of driver cognitive distraction using support vector machines. IEEE Trans. Intell. Transp. Syst. 8 (2), 340–350.
Liu, Y., Wu, C., Zhang, H., Ding, N., Xiao, Y., Zhang, Q., Tian, K., 2024. Safety evaluation and prediction of takeover performance in automated driving considering drivers' cognitive load: A driving simulator study. Transp. Res. Part F: Traffic Psychol. Behaviour 103, 35–52. https://doi.org/10.1016/j.trf.2024.03.021.
Loew, A., Koniakowsky, I., Forster, Y., Naujoks, F., Keinath, A., 2023. The impact of speech-based assistants on the driver's cognitive distraction. Accident Anal. Prevention 179, 106898. https://doi.org/10.1016/j.aap.2022.106898.
Louw, T., Markkula, G., Boer, E., Madigan, R., Carsten, O., Merat, N., 2017. Coming back into the loop: drivers' perceptual-motor performance in critical events after automated driving. Accident Anal. Prevent. 108, 9–18.
Louw, T., Kuo, J., Romano, R., Radhakrishnan, V., Lenné, M.G., Merat, N., 2019. Engaging in ndrts affects drivers' responses and glance patterns after silent automation failures. Transp. Res. Part F: Traffic Psychol. Behaviour 62, 870–882.
Mehler, B., Reimer, B., Coughlin, J.F., Dusek, J.A., 2009. Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. Transp. Res. Rec. 2138 (1), 6–12.
Mehler, B., Reimer, B., Dusek, J.A., 2011. Mit Agelab Delayed Digit Recall Task (n-back), 17. Massachusetts Institute of Technology, Cambridge, MA.
Metz, B., Landau, A., Hargutt, V., 2015. Frequency and impact of hands-free telephoning while driving – results from naturalistic driving data. Transp. Res. Part F: Traffic Psychol. Behaviour 29, 1–13.
Mishler, S., Chen, J., 2024. Boring but demanding: using secondary tasks to counter the driver vigilance decrement for partially automated driving. Hum. Factors 66 (6), 1798–1811.
B. Mok, M. Johns, K.J. Lee, D. Miller, D. Sirkin, P. Ive, W. Ju, Emergency, automation off: Unstructured transition timing for distracted drivers of automated vehicles, in 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 2015, pp. 2458–2464.
Mok, B., Johns, M., Miller, D., Ju, W., 2017. Tunneled in: drivers with active secondary tasks need more time to transition from automation. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17. Association for Computing Machinery, New York, NY, USA, pp. 2840–2844.
Morando, A., Victor, T., Bengler, K., Dozza, M., 2021. Users' response to critical situations in automated driving: Rear-ends, sideswipes, and false warnings. IEEE Trans. Intell. Transp. Syst. 22 (5), 2809–2822.
Moray, N., 1967. Where is capacity limited? a survey and a model. Acta Psychol. (Amst) 27, 84–92. https://doi.org/10.1016/0001-6918(67)90048-0.
Naujoks, F., Purucker, C., Neukum, A., 2016. Secondary task engagement and vehicle automation – comparing the effects of different automation levels in an on-road experiment. Transp. Res. Part F: Traffic Psychol. Behaviour 38, 67–82.
Naujoks, F., Höfling, S., Purucker, C., Zeeb, K., 2018. From partial and high automation to manual driving: Relationship between non-driving related tasks, drowsiness and take-over performance. Accident Anal. Preven. 121, 28–42.
Navon, D., Gopher, D., 1979. On the economy of the human-processing system. Psychol. Rev. 86 (3), 214.
NHTSA, 2012. Visual-manual Nhtsa Driver Distraction Guidelines for in-Vehicle Electronic Devices. National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT), Washington, DC.
Oostwoud Wijdenes, L., Ivry, R.B., Bays, P.M., 2016. Competition between movement plans increases motor variability: evidence of a shared resource for movement planning. J. Neurophysiol. 116 (3), 1295–1303. https://doi.org/10.1152/jn.00113.2016 pMID: 27358315.
Oviedo-Trespalacios, O., Haque, M.M., King, M., Washington, S., 2017. Effects of road infrastructure and traffic complexity in speed adaptation behaviour of distracted drivers. Accident Anal. Prevent. 101, 67–77.
Pakdamanian, E., Sheng, S., Baee, S., Heo, S., Kraus, S., Feng, L., 2021. Deeptake: prediction of driver takeover behavior using multimodal data. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21. Association for Computing Machinery, New York, NY, USA, pp. 1–14.
Pan, H., Payre, W., Gao, Z., Wang, Y., 2024. Exploring driving anger-caused impairment of takeover performance among professional taxi drivers during partially automated driving. Accident Anal. Preven. 205, 107686. https://doi.org/10.1016/j.aap.2024.107686.
Politis, I., Brewster, S., Pollick, F., 2017. Using multimodal displays to signify critical handovers of control to distracted autonomous car drivers. Int. J. Mobile Human Computer Interaction (IJMHCI) 9 (3), 1–16.
Radlmayr, J., Gold, C., Lorenz, L., Farid, M., Bengler, K., 2014. How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. Proc. Human Factors Ergonomics Society Annual Meeting 58 (1), 2063–2067.
Rann, J.C., Almor, A., 2022. Effects of verbal tasks on driving simulator performance. Cogn. Res. Princ. Implic. 7 (1), 12.
Regan, M.A., Lee, J.D., Young, K., 2008. Driver Distraction: Theory, effects, and Mitigation. CRC press.
R.S. Rice, F. Dell'Amico, An experimental study of automobile driver characteristics and capabilities, Tech. rep., Technical Report DTIC Document (1974).
SAE International, 2021. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE International, Warrendale, PA, USA.
Sanghavi, H., Zhang, Y., Jeon, M., 2020. Effects of anger and display urgency on takeover performance in semi-automated vehicles. In: 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '20. Association for Computing Machinery, New York, NY, USA, pp. 48–56.
Sanghavi, H., Zhang, Y., Jeon, M., 2023. Exploring the influence of driver affective state and auditory display urgency on takeover performance in semi-automated vehicles: experiment and modelling. Int. J. Human-Comput. Stud. 171, 102979.
Schwalk, M., Kalogerakis, N., Maier, T., 2015. Driver support by a vibrotactile seat matrix– recognition, adequacy and workload of tactile patterns in take-over scenarios during automated driving. Procedia Manuf. 3, 2466–2473.
Shin, H.S., Jung, S.J., Kim, J.J., Chung, W.Y., 2010. Real time car driver's condition monitoring system. SENSORS 951–954.
Sodnik, J., Dicke, C., Tomažič, S., Billinghurst, M., 2008. A user study of auditory versus visual interfaces for use while driving. In. J. Human-Comput. Studies 66 (5), 318–332. https://doi.org/10.1016/j.ijhcs.2007.11.001.
Solís-Marcos, I., Ahlström, C., Kircher, K., 2018. Performance of an additional task during level 2 automated driving: an on-road study comparing drivers with and without experience with partial automation. Hum. Factors 60 (6), 778–792.

Solhjoo, S., Haigney, M.C., McBee, E., van Merrienboer, J.J.G., Schuwirth, L., Artino, A. R., Battista, A., Ratcliffe, T.A., Lee, H.D., Durning, S.J., 2019. Heart rate and heart rate variability correlate with clinical reasoning performance and self-reported measures of cognitive load. Sci. Rep. 9 (1), 14668.
Solovey, E.T., Zec, M., Garcia Perez, E.A., Reimer, B., Mehler, B., 2014. Classifying driver workload using physiological and driving performance data: Two field studies. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14. Association for Computing Machinery, New York, NY, USA, pp. 4057–4066.
Strayer, D.L., Johnston, W.A., 2001. Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. Psychol. Sci. 12 (6), 462–466. https://doi.org/10.1111/1467-9280.00386 pMID: 11760132.
Strayer, D.L., Turrill, J., Cooper, J.M., Coleman, J.R., Medeiros-Ward, N., Biondi, F., 2015. Assessing cognitive distraction in the automobile. Hum. Factors 57 (8), 1300–1324.
Strayer, D.L., Cooper, J.M., Turrill, J., Coleman, J.R., Hopman, R.J., 2017. The smartphone and the driver's cognitive workload: a comparison of apple, google, and microsoft's intelligent personal assistants. Canadian J. Exper. Psychol. 71 (2), 93–110.
Thompson, C.G., Kim, R.S., Aloe, A.M., Becker, B.J., 2017. Extracting the variance inflation factor and other multicollinearity diagnostics from typical regression results. Basic Appl. Soc. Psych. 39 (2), 81–90.
UNECE, 2021. UN Regulation No. 157 - Automated Lane Keeping Systems (ALKS). United Nations Economic Commission for Europe (UNECE), Geneva, Switzerland, 75-137.
van der Heiden, R.M., Iqbal, S.T., Janssen, C.P., 2017. Priming drivers before handover in semi-autonomous cars. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, CHI '17. Association for Computing Machinery, New York, NY, USA, pp. 392–404.
Vogelpohl, T., Kühn, M., Hummel, T., Vollrath, M., 2019. Asleep at the automated wheel—sleepiness and fatigue during highly automated driving. Accident Anal. Prevent. 126, 70–84, 10th International Conference on Managing Fatigue: Managing Fatigue to Improve Safety, Wellness, and Effectiveness".
Wagner, A.R., Borenstein, J., Howard, A., 2018. Overtrust in the robotic age. Commun. ACM 61 (9), 22–24.
Walch, M., Lange, K., Baumann, M., Weber, M., 2015. Autonomous driving: Investigating the feasibility of car-driver handover assistance. In: Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '15. Association for Computing Machinery, New York, NY, USA, pp. 11–18.
Wan, J., Wu, C., 2018. The effects of lead time of take-over request and nondriving tasks on taking-over control of automated vehicles. IEEE Trans. Hum. Mach. Syst. 48 (6), 582–591.
Wandtner, B., Schömig, N., Schmidt, G., 2018. Effects of non-driving related task modalities on takeover performance in highly automated driving. Hum. Factors 60 (6), 870–881.
Wang, M., Lee, S.C., Montavon, G., Qin, J., Jeon, M., 2022. Conversational voice agents are preferred and lead to better driving performance in conditionally automated vehicles. In: Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, AutomotiveUI '22. Association for Computing Machinery, New York, NY, USA, pp. 86–95.
Wangwiwattana, C., Ding, X., Larson, E.C., 2018. Pupilnet, measuring task evoked pupillary response using commodity rgb tablet cameras: Comparison to mobile, infrared gaze trackers for inferring cognitive load. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 1 (4) jan.
Wickens, C.D., 2002. Multiple resources and performance prediction. Theor. Issues. Ergon. Sci. 3 (2), 159–177. https://doi.org/10.1080/14639220210123806.
Wickens, C.D., 2008. Multiple resources and mental workload. Hum. Factors 50 (3), 449–455.
Yang, S., Kuo, J., Lenné, M.G., 2021. Effects of distraction in on-road level 2 automated driving: Impacts on glance behavior and takeover performance. Hum. Factors 63 (8), 1485–1497. https://doi.org/10.1177/0018720820936793 pMID: 32677848.
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 Anal. Prevent. 92, 230–239. https://doi.org/10.1016/j.aap.2016.04.002.