Crying jaywalker! Notifying take-over-requests and critical events in operational driving domain of autonomous vehicles via multimodal interfaces

Crying Jaywalker! Notifying Take-Over-Requests and Critical Events in Operational Driving Domain of Autonomous Vehicles via Multimodal Interfaces

Filippo Gabriele Pratticò
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy
filippogabriele.prattico@polito.it

Lorenzo Valente
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy
lorenzo.valente@polito.it

Fabrizio Lamberti
Dipartimento di Automatica e Informatica
Politecnico di Torino
Torino, Italy
fabrizio.lamberti@polito.it

Abstract

The advent of self-driving cars promises to enable occupants to repurpose commuting time. However, although in conditional automation (SAE Level 3) drivers can engage in non-driving related tasks (NDRTs), they must be ready to intervene when prompted by the system with a take-over request (TOR). The vehicle may also need to warn the driver about critical events without notifying a TOR (as in sudden hard braking due to a jaywalker). Clearly and effectively communicating these events and their urgency is crucial for the successful adoption of autonomous vehicles.

This work analyzes the impact of multimodal visual and audio cues in conveying this information. It considers an augmented reality (AR) windshield display (WSD) combining screen-fixed elements and world-registered AR overlays, alongside an auditory interface providing explanations and alerts through speech and abstract sounds. The effectiveness of these combined stimuli was evaluated through a user study conducted in a VR-based driving simulator.

CCS Concepts

• **Human-centered computing → Empirical studies in visualization**; *Empirical studies in HCI*; **Interactive systems and tools**.

Keywords

interruption, NDRT, notification, autonomous vehicle, HMI, TOR

ACM Reference Format:

Filippo Gabriele Pratticò, Lorenzo Valente, and Fabrizio Lamberti. 2025. Crying Jaywalker! Notifying Take-Over-Requests and Critical Events in Operational Driving Domain of Autonomous Vehicles via Multimodal Interfaces. In *30th International Conference on Intelligent User Interfaces (IUI '25), March 24–27, 2025, Cagliari, Italy*. ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3708359.3712126

1 Introduction

Autonomous vehicles (AVs) are announced to be the next revolution in everyday mobility and transportation. With the continuous advancements in the enabling technologies of machine learning, simulation, sensing, and computing combined with the competition among car manufacturers driven by the desire to be the first to capture a significant share of this burgeoning market, the prospect of witnessing fully autonomous vehicles on the roads in the next few years is becoming concrete [19, 62]. The deployment of AVs at SAE Level 4 (L4), such as Waymo's self-driving cars [56], demonstrates the feasibility of these systems, albeit with a limited operational design domain[1] (ODD) and in a confined area.

The positive impacts expected from AVs are manifold, from increasing safety by limiting the risks of road accidents and fatalities (which are mainly relatable to human factors [41]) to reducing fuel-consumption, traffic and pollution, which is key for environmental sustainability of transportation [7, 19]. However, these vehicles are – and is speculated will still be for at least the next decade – prohibitively expensive for the average consumer. This fact may indeed hinder their widespread adoption [62]. At the same time, though, AVs classified as SAE Level 3 (L3), also known as conditional automation level, are expected to gain traction in the near future, being considered as a more affordable evolution of the already fairly common SAE Level 2 (L2) AVs.

Unlike L2 AVs, in which the driver is asked to supervise the AV while the system is driving autonomously and intervene if deemed necessary, in L3 AVs the entire dynamic driving task (DDT) is carried out within the ODD without the need for the driver to supervise. The driver is then allowed to comfortably engage in non-driving related tasks (NDRTs), such as eating, reading, listening to music/podcast, watching movies/videos or even sleeping [50, 52], as it is customary for instance when spending transit time in other forms of public transportation. As a matter of fact, the likelihood that drivers engage in such NDRTs is significantly higher with increasing levels of automation since, from an individual's perspective, one of the most compelling reasons for adopting AVs is the opportunity to repurpose the time spent on the road [13]. Nonetheless, in a SAE L3 AV, the driver must still remain available as a "fallback-ready user" who intervenes upon request. Specifically, when the AV approaches its ODD limits or a system failure occurs, it must promptly notify a so-called Take-Over Request (TOR) to the driver who, in turn, is expected to intervene by gaining the full control of the vehicle and responsibility of the DDT within a possibly short timeframe [37, 44].

[1] the sets of conditions and events in which an AV can safely and autonomously execute the dynamic driving task (DDT)

ACM ISBN 979-8-4007-1306-4/25/03
https://doi.org/10.1145/3708359.3712126

This peculiarity of SAE L3 AVs poses significant challenges from the human-machine interaction perspective. In fact, the vehicle's human-machine interface (HMI) has been acknowledged to play a critical role in ensuring that drivers are notified with the appropriate urgency and provided with the information needed to reestablish a sufficient level of situational awareness [14, 33]. There is no surprise then that an extensive body of literature exists in the HMI domain regarding how to deal with TORs, which suggests that multimodal interfaces, particularly those combining audio and visual information, are the most effective way to ensure required urgency and clarity [13, 14, 26]. In terms of reducing the time to re-establish situational awareness, it also appears that interfaces which deliver visual information by leveraging an augmented reality (AR) windshield display (AR-WSD) are more effective than, e.g., head-down dashboards, since the driver can receive AV's perception information and look at the road at the same time [13, 14].

TOR notifications are not the only information that shall be provided to the driver during a ride in a SAE L3 AV. In fact, there is consensus about the fact that, while autonomous driving is engaged, the AV should constantly inform its occupants about its capabilities, status, environment understanding as well as upcoming manoeuvres and TORs (if it is possible to anticipate them, e.g., before encountering an out-of-ODD condition) [31, 33].

However, although notification of TORs is a well explored topic being a safety critical matter, there is a paucity of studies that investigated the other continuous information that shall be delivered in a SAE L3 context. Undeniably, HMIs tailored to this information have been subject to many investigations in works that dealt with higher levels of automation, therefore lacking the need to notify TORs and provide information with potentially conflicting communication channels [31, 58]. These concerns are even more severe by considering the fact that SAE L3 AVs will need to handle autonomously, i.e., within their ODD, even critical events which may require sudden manoeuvres like hard braking, e.g., to deal with an unexpected obstacle or a jaywalker. Providing timely and informed notification of these critical events is crucial for two key aspects: to preserve trust in system's ability to carry out the DDT [14, 33] and to let passengers anticipate inertial effects (accelerations), which is of primary importance to prevent the onset of self-driving car motion sickness [58]. Furthermore, since the need to notify drivers of critical events alongside TORs can potentially introduce confusion, an HMI being able to differentiate between these notifications in terms of urgency and clarity is crucial to prevent them from misinterpreting the nature of the specific event [33]. In fact, consequences could be either safety critical, i.e., the driver mistakenly takes control during a demanding manoeuvre without being prepared for it, or detrimental for NDRT performance and user experience, forcing unnecessary self-interruptions from the NDRT [20].

With the aim to investigate such context, in this work four designs of unimodal and multimodal notification methods combining visual and auditory channels are proposed and evaluated against each other as well as against a baseline condition using a virtual reality (VR)-based simulation. The objective is to determine which of them, if any, is more appropriate to notify a critical in-ODD event alongside multimodally notified TORs, and if critical events notification may affect the perceived TOR clarity and urgency.

2 Related Work

There is a growing body of literature that focuses on developing ways to enhance the overall driving experience on SAE L3 AVs, aiming to increase trust by improving drivers' awareness as well as providing appropriate information about ongoing events [39] and effective notifications regarding TORs [38].

2.1 Implications of Multitasking in AVs

One of the first factors that comes into consideration when dealing with human factors in SAE L3 AVs are consequences to the scarce human ability to cope with multitasking. This is especially true in environments of dynamic nature like AVs, since drivers must seamlessly transition between primary and secondary tasks while maintaining situational awareness and decision-making capabilities. Existing literature suggests that, while an efficient working memory is a reliable predictor of general multitasking ability, different cognitive faculties may be necessary to rapidly respond to changing task demands in AV-related scenarios and many factors can affect context switching time [40]. Janssen & Kenemans [21], highlighted the complex interplay between human cognition and vehicle automation. While AVs can potentially free cognitive resources for NDRTs, concerns arise about drivers' ability to regain control when necessary.

A framework based on situation awareness has been proposed by Skrypchuk et al. [51], which aims to address the challenges of managing multiple, unrelated tasks simultaneously in vehicles and inform the design of in-vehicle information systems. The authors postulated that drivers develop separate bodies of knowledge for each of their active goals, and the process of achieving situation awareness is expanded to incorporate this concurrent development of goal-directed knowledge. In line with this framework there is the "threaded cognition" model by Salvucci er al. [46], which offers an integrated theory of concurrent multitasking, explaining how multiple streams of thought can be coordinated and executed across available cognitive resources. Effects of multitasking on SAE L3 AVs with respect to TORs perfomance have been specifically studied by Li et. al. [29], who observed prolonged takeover time and slowed decision-making under multitasking circumstances. The introduction of multimodal interfaces as a way to mitigate potential multitasking inefficiency is supported by Detjen et al. [6], who found also selectivity for the sensory channel involved in the multitasking, indicating that the kind of secondary task or NDRT may lead to specific and asymmetric delays in recovering input control over channels involved in the tasks.

2.2 Interfaces in L3 AVs

Over the years, various HMIs have been proposed to enhance trust in the vehicle and provide the driver with supportive information while potentially limiting the burden of information overload. Literature works initially focused on utilizing systems already present in non-autonomous vehicles to convey vehicle's status in the best way possible leveraging available channels [39].

2.2.1 Visual Channel. It is now well established from a number of previous works that HMIs exploiting the visual channel to display elements and cues can be classified into: (i) dashboards, namely, behind the steering wheel and/or infotainment central dashboards; (ii) nomadic devices, such as tablets or mobile devices which, despite being fairly popular in studies dealing with SAE L4/5 AVs, are not typically considered in SAE L3 scenarios since they pose an additional barrier when the driver needs to regain control of the steering wheel being his or her hands busy; (iii) vehicle interior interfaces, such as environmental light stimuli projected in the cabin, which are especially employed with the aim to mitigate motion sickness [43, 58, 59], and AR head-up displays (HUDs), which visualize synthetic visual cues on a small size AR-HUD or on the full windshield area as an overlay of the outside view.

Regarding early studies that explored how to convey information using the visual channel, Forster et al. [11] studied how to display the vehicle's system status directly on the dashboard. By designing HMIs for the dashboard, drivers could access information like current driving state, speed limits, and upcoming manoeuvres. Testing with a low-fidelity non-immersive VR simulator demonstrated a clearer driving experience and safer interaction between the driver and the automation. Jose et al. [23] analyzed the impact of AR techniques to increase clarity during car navigation, comparing an AR-WSD with a dashboard in a virtual simulation. The study found that the AR-WSD improved driving performance by reducing errors due to better visibility and ease of use, and it was preferred over the other interface.

Building on the enhanced clarity of information visualized on WSDs, previous research has explored where and how to present information most efficiently on these displays. Haeuslschmid et al. [15] investigated user preferences for positioning various useful information, such as urgent warnings, vehicle data, and textual content, on the WSD. Their findings showed that most users preferred the information that has to be shown continuously to be displayed in the left area in front of the driver, avoiding the central field of view (FOV). Following these recommendations, Lindemann et al. [30] proposed an AR-HUD design aimed to reduce driver discomfort from unpredictable hazardous situations in urban traffic environments, leading to an increase in situational awareness in case of critical events. The proposed design featured a destination distance and time panel at the top, a navigation panel linked to speed and upcoming directions at the bottom, and a traffic panel in the center that was activated only during slow driving conditions.

2.2.2 Auditory Channel. Previous research classified auditory interfaces into two categories based on cue type: speech-based and non-speech-based. Speech-based interfaces are further divided in those leveraging speech, i.e., short fragments of utterance prerecorded or generated using text-to-speech technology to deliver information with a human-like voice, and those relying on spearcons, i.e., speech accelerated by 40-50%, aimed to deliver the same information in a shorter time compared to speech while preserving comprehensibility [5]. Non-speech-based interfaces, in turn, are divided into those using abstract synthetic sound (also known as earcons), whose encoding of message urgency has been extensively studied in human factors literature by analyzing several sound features and parameters [28], and those exploiting auditory icons, i.e., natural sound that conveys meaning or information through its resemblance to a real-world sound or event (e.g., the sound of a car crash to notify a road accident).

Concerning initial studies that explored how to convey information using these interfaces, Forster et al. [12] explored how speech notifications impact driver trust and acceptance of the AV as well as clarity of communication. Their study, based on a VR simulation, focused on speech descriptions of events typically managed by SAE L3 autonomous driving that may alert or heighten driver anxiety due to anticipated decelerations, such as speed limits and lane changes. Experimental results showed that speech notifications enhanced trust, anthropomorphism, preference, situational awareness, and alertness, making users perceive the autonomous driving system as more anthropomorphic. Building on the clarity provided by speech descriptions, Zhang et al. [61] focused on the content of the speech and the timing of the communication. Their findings indicate that "what" explanations, namely describing what is happening, can increase anxiety by making the action appear as more dangerous and urgent, while "why" explanations, namely providing the reasoning behind the action, and "what + why" explanations enhanced clarity and sense of trust in the AV. Additionally, providing explanations before actions were actually performed by the vehicle reduced confusion.

As anticipated, the body of literature on human factors and HMI has identified different types of audio interfaces that can be provided using the audio channel. In the context of SAE L3 AVs, Wright et al. [57] investigated how different audio types affect drivers' situational awareness in hazard anticipation scenarios, comparing the impacts of earcons, speech, and their combination on drivers' reactions before entering potential hazard zones. Results obtained through a VR driving simulation showed that earcons significantly improved reaction time, while speech enhanced comprehension. However, the combination of earcons and speech did not yield significant changes.

Previous studies indicate that providing a speech explanation before an action is performed by the AV is linked to increased positive attitudes (trust, anthropomorphism, preference, situational awareness, and alertness) and reduced negative emotions. Regarding content, the "what"-only explanation resulted in the poorest performance and lowest acceptance compared to the "what + why", "why"-only, and "no-explanation" conditions. Both "why"-only and "what + why" explanations showed positive outcomes in trust and acceptance. It should be also outlined that variations in the effectiveness of presenting signals through visual means versus auditory means have been observed. For instance, auditory signals have generally been proven more effective than visual signals in conveying hazard alerts quickly and helping to understand the magnitude of potential hazards [4]. However, it is worth noting that the auditory modality has also been linked to heightened levels of annoyance when compared to the visual modality [32]; hence, it is still disputed what of these two unimodal notifications shall be preferred, especially when multiple notifications with different urgency levels have to be prompted.

2.3 TOR Notification

Researchers have built on the knowledge extracted about conveying supportive information to drivers to deal with the design of interfaces for TOR notification capable of promptly re-engaging drivers when factors that necessitate human intervention arise [45].

For instance, Chai et al. [5] investigated the effect of auditory interfaces targeted to TOR notification on driver performance, focusing on the potential of different design choices in a VR simulated driving scenario. The results showed that non-speech-based interfaces were the most effective for alerting drivers to re-enter the control loop, while speech-based interfaces were appreciated for their ease of understanding, though they led to relatively poorer TOR performance. More specifically, within the same category, speech elicited faster reactions than spearcons, and earcons outperformed auditory icons.

Regarding visual interfaces, Karatas et al. [24] evaluated the effectiveness of an AR-HUD compared to a static HUD (S-HUD) during a TOR triggered by a jaywalker. The study focused on determining whether the two HUDs could reduce driver reaction time and increase situational awareness to avoid potential collisions with pedestrians through a driving simulator experiment. According to experimental results, the use of an AR-HUD to signal a TOR led to a significantly faster recognition of jaywalking pedestrians and improved driver response, enhancing situational awareness and overall trust. However, subjective feedback suggested the incorporation of multimodal alerts, such as auditory cues, alongside visual ones to further increase understanding and situational awareness.

2.4 Multimodal Notifications

An example of multimodal interface for TOR notification is provided in [25], where Kim et al. analyzed the impact of a visual TOR and two types of auditory TORs (auditory icons and earcons, respectively) in providing manoeuvre information. Focusing on trust, situational awareness, and sense of control and leveraging a driving simulator with a full cockpit setup, they found that the auditory TOR significantly improved user experience by enhancing all the metrics compared to the visual TOR. The effectiveness of the auditory TOR varied depending on the manoeuvre: the combination of auditory icons and visual TOR outperformed the other unimodal alternatives in terms of situational awareness for overtaking and roundabouts, while earcons performed better for lane changes. Similarly, Yoon et al. [60] explored how different TOR modalities (visual, auditory, vibrotactile) could influence drivers' responses when engaged in NDRTs. Their study specifically focused on analyzing drivers' reaction times when re-entering the control loop after a TOR, while also assessing perceived safety and effectiveness of the various modalities. Results obtained in a driving simulator showed that visual-only TORs resulted in longer response times and lower safety perception, according to previous findings. In contrast, multimodal TORs incorporating auditory alerts performed better, significantly improving both drivers' response times and perceived safety.

Drawing on the insights gained for multimodal TORs, Huang et al. [17] evaluated the effects of speech-based auditory, visual AR-HUD, and multimodal TORs on drivers' behavior through an immersive driving simulation in VR. To ensure a fair comparison of the various TOR modalities, they included a baseline condition with no interface. Focusing on TOR triggering events, their study exploited the sudden appearance of jaywalking hazards, such as animals, pedestrians, or other vehicles, while using the same TOR notification for events of varying severity. The experiment was carried out using a virtual environment in which the driver interacted with the simulation using an head-mounted display (HMD), a steering wheel, and pedals. The study metrics focused on analyzing drivers' reaction times, anxiety levels, perceived trust, and preferred control methods. Regarding reaction times, in contrast to previous study [60], results showed that the AR-HUD visual TOR reduced reaction times, while the speech-based auditory TOR did not improve them but enhanced situational awareness. In line with earlier findings, the multimodal TOR showed an overall improvement in reaction times. Concerning anxiety and trust, the results showed an improvement with both unimodal TORs and an additional enhancement with multimodal TOR, which also aligned with preferences expressed regarding the various TOR modalities.

Although the effectiveness of multimodal TORs in managing critical events has been confirmed as much as the importance of providing continuous feedback to the driver [33, 47], there is a paucity of studies that evaluated the coexistence of these stimuli. More importantly, to the best of the authors' knowledge, there is no study that analyzed the impact and potential conflict among in-ODD and out-of-ODD critical event notifications.

The objective of this work is to determine which unimodal or multimodal notification method is the most appropriate for providing information about critical in-ODD events alongside with multimodal TOR notifications, while fostering trust in the system without being overly intrusive or causing undue stress, such as confusion from unclear messages. It is crucial to ensure that drivers are not misled into thinking they need to intervene unnecessarily, as the autonomous system can manage such situations. Ideally, the notification of in-ODD critical events should aim to convey the message to the driver whilst preventing him or her from interrupting or distracting from their NDRTs and preserving the urgency of the TOR notification.

3 Material and Methods

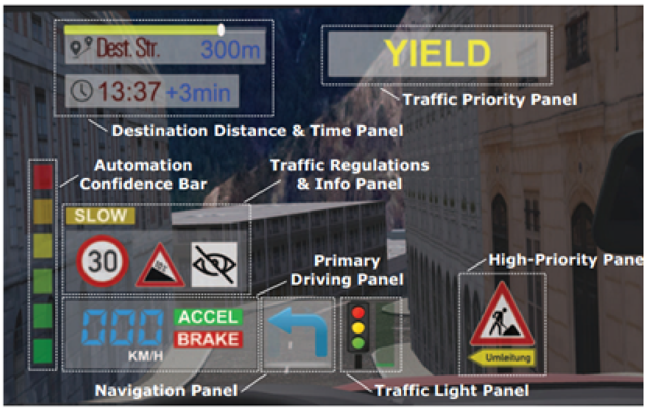
This section reports on the proposed notification variants and provides details about the VR-based simulation system that was devised in order to perform the experimental evaluation under safe and repeatable conditions.

3.1 Critical Event Notification Variants

Based on the analysis of the previous literature on notification interfaces (Section 2), it appears that the most promising ones are those based on speech (S) that leverage the "what+why" formula. Specifically, to notify the jaywalker (chosen as the in-ODD critical event in this study as later explained in Section 3.2.2) the speech message implemented was *"Braking! Jaywalker detected"*.

The unimodal visual interface selected was instead based on a 3D AR bounding box displayed via an AR-WSD [5, 31] (later shortly referred to as V). This visual interface highlights the jaywalker throughout the whole crossing time by enclosing it in a red-colored world-contextualized box-shaped visual element, as depicted in Figure 1c.

The combination of these two interfaces was selected as multimodal interface and included in the evaluation, referring to it as visual + speech (VS).



(a) AR-WSD HMI as presented in [30]

(b) AR-WSD HMI, dashboard and infotainment display devised in this work

(c) 3D AR bounding box used to notify the jaywalker via visual cue

(d) HMI during the early stage of a TOR

Figure 1: HMI of the AV devised for the study.

Table 1: Summary of the proposed variants to notify the in-ODD critical events (i.e., jaywalkers).

Configuration	Baseline	Visual	Speech	Visual + Speech	Visual + Abstract Sound
Notification modality	/	3D AR Bounding Box	Speech (What + Why)	3D AR Bounding Box + Speech (What + Why)	3D AR Bounding Box + Abstract Sound

In addition, whilst abstract sounds have been observed to be less effective than speech or spearcons for the envisaged notification context [5], they have been found promising when combined with visual stimuli [17, 60]; hence, a visual+abstract (VA) sound based interface was included in the study, whereas the abstract sound only was not considered. The abstract sound was crafted following the guidelines in [28, 35], characterizing it with a different and less hazardous sound than the one used to notify the TOR (Section 3.2.3).

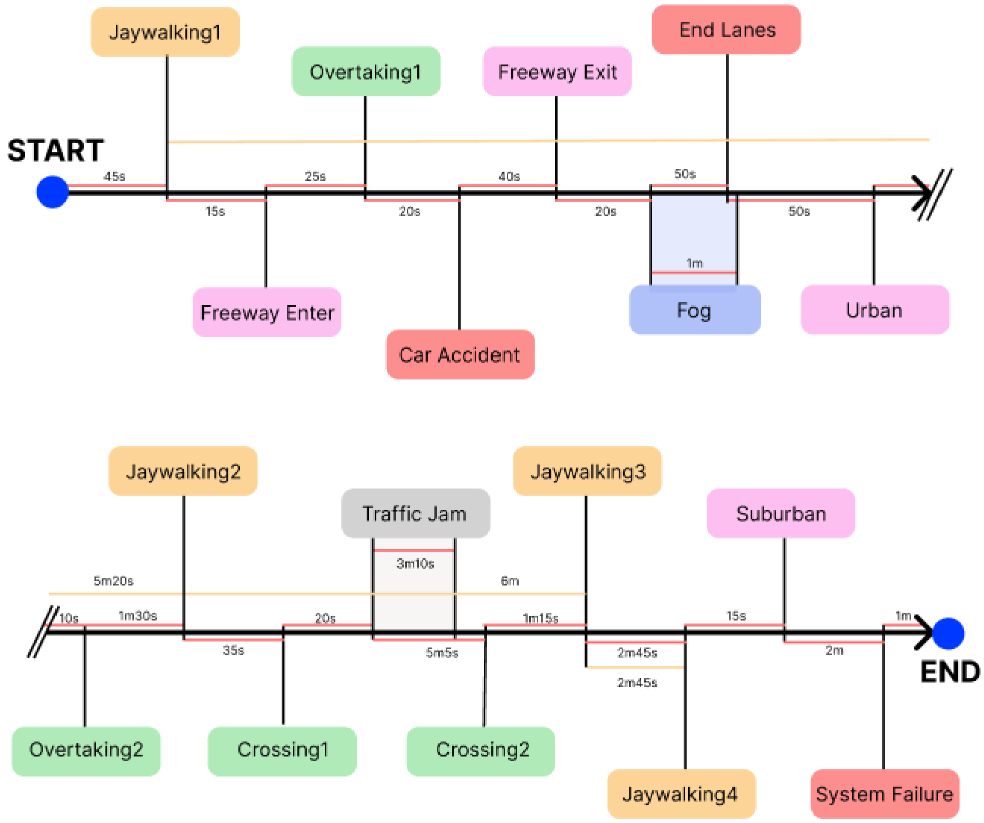
Table 1 summarizes the variants proposed in this work to be evaluated against each other and a baseline condition.

3.2 Simulation System

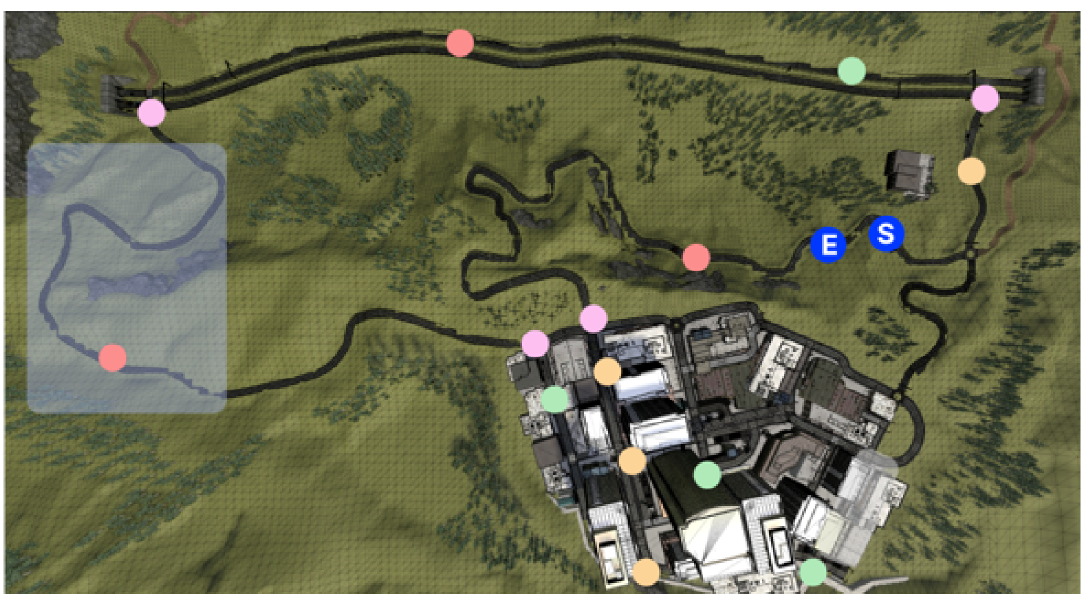
The VR-based system that was arranged for the experiments simulates a movie watching NDRT via a head-down display of the infotainment system during a ride on a SAE L3 AV.

3.2.1 Hardware & Configuration. A configuration similar to the one adopted in the work by Ihemedu-Steinke et al. [18] was devised, based on an HMD for immersive VR in combination with a motion simulation platform.

The HTC Vive Pro [55] was used as HMD. This HMD features a resolution of 1400×1600 pixels per eye, spanning a horizontal FOV of 110° at a 90Hz refresh rate. To enable six degrees-of-freedom tracking, the Vive Lighthouse v2.0 [55] technology was chosen, with two infrared laser emitters (a.k.a. base stations) installed in the room.



(a) Defined journey with events and time spacing: manoeuvres domain transition degraded driving condition traffic status update in-ODD critical event TOR



(b) Devised ride path and location of events in the Windridge city map

Figure 2: Devised ride journey and scenario.

The Atomic Motion Systems A3 was employed as motion simulator (analogously to [18]). It consists of an adjustable racing seat which can accommodate individuals with heights ranging from 1.35m to 2.00m mounted on a two degrees-of-freedom tilting mobile platform with DC motors actuators and the pivot point underneath the seat. The motion simulator operates with an update rate up to 100Hz and supports a maximum speed of 71.3°/s and a movement range of 27°, resulting into maximum exerted accelerations of $4.5m/s^2$ which allow rendering a fair portion of the dynamics that could be experienced during a car ride [3, 22]. The motion simulator was interfaced via the Actuate Motion [1] proprietary software and driver, configured with the built-in communication protocol based on shared memory and coupled with the respective plugin to share acceleration data with the Unity-based simulation environment (Section 3.2.2).

A Vive tracker (v3.0) [53] was mounted on the seat to enable motion compensation, which was applied using the OpenVR Motion Compensation plugin of SteamVR compositor software [36]. Motion compensation is required to refer the movements of tracked objects (i.e., the HMD) as to be relative to the seat (hence to the simulated AV) rather than to the actual room in which the motion platform is located. It works both by applying geometrical transformations and by cancelling out the accelerations due to the motion platform, since the Vive Lighthouse is an hybrid tracking technology that uses both optical inside-out active constellation matching and inertial sensing.

The user was enabled to manually drive the vehicle using a Fanatec steering wheel and pedals kit (ClubSport Wheel Base V2.5 and ClubSport Pedals V3) [9]. One of the built-in buttons on the steering wheel was used to engage the autonomous driving mode. To disengage autonomous driving, it was decided to detect user inputs on the brake pedal (threshold at 5% of braking power) and steering wheel (threshold was set at 3° of angular deviation from the simulated steering wheel). The steering wheel is capable of providing high-fidelity force feedback via brushless servo motor with dual v-belts. The force feedback has been also used to control the steering wheel angle during the simulated autonomous drive, moving it for synchronizing the real and virtual steering wheel using a PID controller module.

The simulation was run on a Windows 10 computer with Intel Core i9-12900K processor, 64GB DDR5 4GHz RAM and a NVidia GeForce RTX 4080 graphics processor.

3.2.2 Simulated Environment & Scenario. The simulation environment was implemented in the Unity game engine (v2020.3 LTS) [54] and built upon the CiThrus open-source project [34]. CiThrus, in turn, is built on top of the Unity AirSim open-source software [49] (formerly developed to generate synthetic datasets for training AVs' computer vision algorithms) and the "Windridge City" scenario, and adds to them pedestrian simulation as well as a waypoint-based deterministic traffic simulation.

The simulation environment was further optimised for an effective fruition via immersive VR devices with the aim of increasing visual fidelity and minimising the potential onset of cybersickness by targeting a constant framerate of 90fps. In particular, the implementation of traffic collision avoidance was reworked using Unity's Data-Oriented Technology Stack (DOTS) to optimize computations, since the traffic simulation consists of multiple instances of the same object (cars) that can follow the same logic. Furthermore, collision detection was made less resource intensive by simplifying mesh colliders and parallelizing it leveraging Unity's "collision queries" feature. Finally, the graphics were revamped using Unity's HDRP-VR (High Definition Render Pipeline for VR) pipeline, traffic vehicles and pedestrians were rendered using GPU instancing, and native data collection from the participant AV's simulated sensors (for computer vision training) was disabled.

Besides the above improvements, the waypoint-based system was endowed with the possibility to set the speed at which a vehicle has to go through each waypoint since, in the proposed implementation, this information is also leveraged by the module in charge of the participant AV's drive simulation (Section 3.2.3). The "Windridge City" scenario was also adapted, translating the US road signalling system and conventions to EU ones, since the participants of the experiment were anticipated to belong to the Italian population (Section 4.1).

Journey. A journey was devised so that it encompasses a ride along the highway, the sub-urban and the urban-downtown areas of the map. The journey also included a set of relevant events extracted from the literature[5, 27, 31, 42]. Specifically, events have been classified into the following categories:

- TORs (roadworks or car accidents) and degraded driving conditions (which lead to a TOR);
- Manoeuvres (overtaking or stopping at a regulated pedestrian crossing, in urban area);
- Critical in-ODD (a jaywalker);
- Domain transitions (ingress and egress to/from freeway, sub-urban or urban area);
- Traffic status updates (traffic jam).

The journey included at least one event for each category. Moreover, for the two categories of particular interest for the study, three TORs and four jaywalking events (two crossing from left-to right, the other two the other way) have been implemented. The jaywalker has been selected, among others, as a critical in-ODD event since expected to stimulate a high intrinsic sense of urgency and emotional response in the driver [24, 42].

The journey ride, with an approximate total ride time of about 20 minutes, together with events location in the Windridge City map and timing are depicted in Figure 2.

To account for potential variability due to the time spent in manual driving, the implementation of the events followed a trigger based logic. Specifically when the AV passes by some predefined locations in the virtual scenario, a choreography is triggered by leveraging the Unity timeline tool. This tool is capable of both triggering callback functions (required to have the HMI react accordingly) or initiating predefined animations, such as in the case of the jaywalker.

3.2.3 Simulated AV. Whilst the traffic vehicles were implemented with computational efficiency in mind, the participant's AV is conceived to prioritize the fidelity of the dynamic simulation of the car ride. In particular, the NWH Vehicle Physics 2 plugin was used to simulate the vehicle's physics. This plugin provides a modular architecture with customizable components including suspensions, powertrain, wheel & tire mechanics, and supports realistic simulation of inertia, torque, engine behavior, and aerodynamics, making it adaptable to a wide range of vehicles. The system was configured with an electric engine powertrain and sedan car dynamics.

The autonomous drive simulation exploits the *autonomous module* of NWH (based on PID controllers), which was customized to make it compatible with the path implementation and drive the vehicle accordingly.

As the intention was to seat the participant in the spot usually reserved for the driver (i.e., the front-left one, based on the target population), the vehicle 3D model and interior design of the Unity automotive HMI template project [16] was used as a base.

AV Base HMI. The base HMI layout in which critical events and TOR notifications will be prompted is grounded on previous work from the literature and actual implementations, being made of an AR-WSD, a dashboard, and an infotainment central dashboard display. In particular, the informative dashboard is inspired to that of Tesla Model X and provides a synthetic view of the AV's perception, a map-based navigator, some information about vehicle's status (e.g., speed, battery level) and, while in manual drive, whether the autonomous mode is available. The central infotainment display is 15" with a 16:9 aspect ratio. The HMI is completed by the audio playback system, which was implemented with spatialized audio as if the sounds (i.e., abstract sounds and speech) were sourced from the dashboard location.

The AR-WSD follows the design by Lindemann et al. [30]. The AR-WSD display shows both screen-fixed elements as well as world-registered AR overlays. The screen-fixed elements are organized into various sections. The driving panel is located on the bottom-left corner of the windshield and provides information about the current speed and AV accelerating/braking inputs with two separate elements. Directly above the driving panel is positioned the traffic regulation panel, which serves the purpose of presenting information such as the current speed limit and echoing traffic signs. In the case multiple signs are active simultaneously, the position of the pictograms is adjusted to fit them into a visual array. Located on the top-left area of the windshield there is the time and destination panel, which displays crucial information including the current time, estimated time of arrival, and distance to the destination. The navigation panel, located on the right of the driving panel, presents information that anticipate the manoeuvrers regarding the upcoming turn directions through the use of specialized pictograms for each direction. Alongside is placed a traffic light area. On the left side of the driving panel there is a 6-level confidence bar, which provides a visual representation of the autonomous DDT reliability confidence level (throughout the simulated journey, the value of this bar changed just when encountering the foggy road and when the system failure occurred). Positioned at the bottom-central part of the windshield, the high-priority panel showcases critical information, such as a blinking pictogram whenever a TOR is notified. The pictogram carries the information regarding the occurrence of the TOR and the underlying reason for it.

All the AR-WSD elements were implemented using the Unity's UI system. Specifically, a UI canvas was used to define the AR-WSD area on a portion of the windhield 3D mesh (that has transparent material). TextMeshPro with appropriate fonts was used to display and update all textual information in real-time, whereas all the other picograms and symbols were prepared as Sprite UI elements and orchestrated by scripts. In addition, in order to convey the feeling of AR elements (mimicking display technologies like birdbath or waveguide optics which are expected to be used in real-world implementations), the visual density of the AR-WSD was reduced by capping the maximum opacity at 80%.

The remaining part of the AR-WSD is used to show the world-registered AR overlays. 2D bounding boxes are used to highlight both the encountered traffic signs and traffic lights. The 2D boxes consists of fully transparent rectangles with a dashed style border enclosing the element of interest. Different colors are used for the border based on the level of hazard (e.g., red for red traffic lights and roadworks signs, yellow for speed changes and yield traffic signs). The rectangles are implemented as 3D meshes (planes with custom shaders) in the virtual environment and their activation is orchestrated via scripts. To emulate the windshield projection, the planes are constantly adjusted as to be parallel to the virtual camera plane, and custom shaders with hold-out pass used to make the AR overlay visible just if looked through the windshield only. It is worth nothing that if the observer looks at an element which has an AR overlay but through mirrors or side-windows, these will be masked out and not shown. A similar implementation was used for the jaywalker notification that, as anticipated in Section 3.1, was identified by a red-colored 3D bounding box (excepted in baseline and speech only conditions where visual feedback is not provided). Differently than for traffic lights and signs, a 3D rectangular prism shape mesh was used (instead of a plane). The border of the mesh is opaque, whereas the faces are set with 70% transparency (Figure 1c). Finally, orientation of the bounding box follows that of the jaywalking pedestrian (aligned with its movement direction).

The resulting interior and HMI can be seen in Figure 1b.

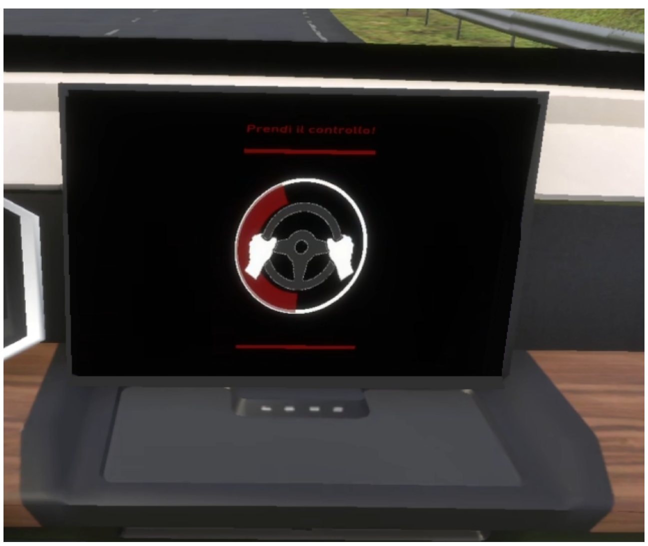
NDRT. The NDRT chosen for the experiment was a movie watching activity [26]. The rationale behind that was to select a hands-free NDRT which involves both visual and auditory sensory channels. For the experiment, a total of 5 different clips (same number of the evaluated conditions) were selected from the Italian TEDx about magic-related topics, with a length approximately matching the expected journey time. During the user study, the clip chosen to play was counterbalanced with respect to the condition evaluated using Latin-square order.

The video is automatically reproduced once the autonomous drive is engaged by the driver; it will pause in case of a TOR notification, whereas in case of the in-ODD critical event notification, will still play. For conditions using the audio channel (S, VS, and VA) the volume of the video is reduced to 10% when the notification is prompted and restored once it ends.

TOR Notification. The implementation of the TOR notification followed the design in [11], and is completed by a "what+why" speech and a bi-tonal abstract sound (designed as urgent following the guidelines in [12, 28]) that is played till the TOR is over.

When an event causing a TOR occurs, the AV begins to decelerate, triggering the AR-WSD as previously described. Simultaneously, the video playback on the infotainment display is paused and an orange circular progress bar appears on a black background alongside a text message and a pictogram, all aimed at notifying the user to

(a) Initial alert phase. Text message, in Italian, translates as "Please. Take over control."



(b) Maximum urgency phase. Text message, in Italian, translates as "Take over control!".

Figure 3: Detail of the two phases of the TOR notification shown on the infotainment display.

take control of the vehicle. As the progress bar reaches its halfway point, both the bar and the accompanying text message change color from orange to red, emphasizing the urgency. If the driver resumes control and switches to manual driving, the central display restores the video playback but overlays a pause icon on the video. Conversely, if the driver does not intervene, the AV will come to a full stop in about 5 to 10 seconds after the TOR notification began.

Videos demonstrating the jaywalker notification with all the proposed variants and the TOR notification are available as supplementary material.

4 Experiment

This section presents the user study performed to evaluate the proposed notification variants, which followed a within-subjects design.

4.1 Sample

The minimum sample size was a-priori estimated as n=24 using G*Power software (v.3.1.9.7) [10] with parameters set: $\alpha = 0.05$, power $(1 - \beta) = 0.99$, Cohen's $f = 0.333(\eta^2 \geq 0.1)$. Participants were recruited through authors' networks of contacts and voluntary word-of-mouth from the subjects involved in the study. Interested subjects were asked to schedule the session in advance using an online booking tool which incorporated also a demographic questionnaire. Subjects with less than three years of driving license, having hearing impairment, or suffering from color blindness were not allowed to participate in the study.

The sample was made of Italian mother tongue subjects, 9 females and 15 males, aged between 21 and 58 y.o. ($\bar{x} = 27.5 \pm 7.1$). They had $\bar{x} = 9.6 \pm 6.7$ years of driving experience, 71% were moderately to very familiar with the use of immersive VR technology or driving simulators, whereas 29% were little to no familiar with them. All the participants reported moderately to high trust in automation and 87% expressed their willingness to ride a real AV.

4.2 Protocol

The participants were asked not to consume alcoholic or psychotropic substances in the 12 hours preceding the experimental session. After signing the informed consent and being instructed on how to operate the emergency stop button on the motion platform, they were let to experience a familiarization ride of about 5 minutes. This ride was arranged in the same scenario (Windridge city) but with traffic and pedestrians (including jaywalkers) deactivated, and participants were instructed how to engage and disengage the autonomous drive, acquaint themselves with the HMI, and observe how a TOR (manually triggered by the administrator) is communicated. In addition, this ride allowed the participants to acclimatize to the simulated environment and scenario, and reduce their curiosity and the so-called "fear of missing out" that could have constituted a confounding factor in the study. In fact, the participant could have spent the ride with the first assigned condition looking outside, hence neglecting the NDRT.

Afterward, they underwent five rides, one for each variant plus an additional baseline (B) condition (i.e., no notification of the jaywalker). The order of exposition was counterbalanced using Latin-square order. The average duration of the experimental session was about 2h30min.

At the end of each ride, the participant was assisted in removing the HMD and completed the post-session questionnaire, which also provided an opportunity to rest; the final questionnaire was administered after the last ride.

4.3 Measures

Subjective feedback was collected by means of a multi-section questionnaire administered after experiencing each condition (29 items), which was made as follows.

Trust. Trust in the AV autonomous drive was measured using 7-point Likert scales from the work by Du et al. [8]. The questionnaire comprises six components: competence, predictability, dependability, responsibility, reliability, faith, plus an overall trust score.

User Experience of Critical Event Notification. User experience factors were measured using the modular version of the user experience questionnaire (UEQ+) [48]. More specifically the compound score for the UEQ+ was computed based on the components of trustworthiness of content, quality of content, and value (14 items in total).

Notification Appropriateness. The critical events notification was also investigated from the aspects of perceived urgency, perceived annoyance, alerting effectiveness, and message clarity by using the four items 7-point Likert scales adapted from [2].

In addition, the same set of aspects were also investigated separately for the TOR notification. Even though the TOR notification modality was kept constant for all the conditions, this measure enabled to observe whether there was any indirect influence or conflicts introduced by the critical event notification.

Ranking (Overall Preference). As said, at the end of the experience, the participants were administered with the final questionnaire in which they were asked to explicitly rank the four variants and the baseline based on their overall preference (1 item).

5 Results

Raw data collected are available as supplemental material.

5.1 Data Analysis

Inferential statistics were computed by using the Friedman's test with Conover correction and the Wilcoxon signed-rank test as post-hoc via MS Excel with the Real-Statistics add-on (v9.3). In the box-and-whisker plots, descriptive statistics are reported as follows: white dot for mean, red dashed line for median, and orange large lines indicators for 95% confidence intervals. Inferential statistics are provided alongside with effect sizes (computed as Cohen's d). In the tables, p-values and relative effect sizes are reported using grayscale coloring (the more significant the darker); in addition, significant differences with p-value ≤ 0.05 are bolded.

5.2 Trust

Regarding Trust (Table 2), significant differences were found for degree of faith, predictability, and overall trust components, whereas for the others the participants reported non-significantly different but fairly high scores.

Specifically, regarding degree of faith, post-hoc comparisons showed a small yet significant advantage of VS against B, and marginally significant differences for VA and V against B. Regarding predictability, post-hoc comparisons showed a moderately significant advantage of VS against B and a small yet significant advantage of VS against V. Furthermore, post-hoc comparisons showed a small yet marginally significant advantage of VA against B and marginally significant advantages of VA and S against V. Regarding overall trust, post-hoc comparisons showed a moderately significant advantage of VS against B.

5.3 User Experience of Critical Event Notification

Considering UEQ+ results (Table 3), significant differences were found for trustworthiness of content, quality of content and value.

Regarding trustworthiness of content, post-hoc comparisons showed substantial significant advantages for all the variants against B, a substantial significant advantage for VS against V, a moderately significant advantage for VA against V, a small yet significant advantage for S against V, and moderately significant advantages for VS against VA and S.

Regarding quality of content, post-hoc comparisons showed substantial significant advantages for all the variants against B, moderately substantial advantages for VS and VA against V, a moderately significant advantage for VS against S, and a small yet significant advantage for VS against VA. Furthermore, post-hoc comparisons showed a small yet marginally significant advantage for VA against S.

Regarding value, post-hoc comparisons showed substantial significant advantages for all the variants against B, a substantial significant advantage for VS against V, and moderately significant advantages for VS against S and VA. Furthermore, post-hoc comparisons showed a small yet marginally significant advantage for VA against V.

5.4 Notification Appropriateness

5.4.1 Critical in-ODD events. Regarding jaywalking notifications (Table 4), significant differences were found for message clarity, perceived urgency, perceived annoyance, and alerting effectiveness.

Message Clarity. Regarding message clarity, post-hoc comparisons showed substantial significant advantages for all the variants against B, substantial significant advantages for VS and S against V, a moderately significant advantage for VA against V, and substantial significant advantages for VS against S and VA.

Perceived Urgency. Regarding perceived urgency, post-hoc comparisons showed substantial significant advantages for VS, VA and S against B, substantial significant advantages for VS and VA against V, a moderately significant advantage for S against V, and a moderately significant advantage for VS against S. Furthermore, post-hoc comparisons showed a moderately marginally significant advantage for V against B.

Perceived Annoyance. Regarding perceived annoyance, post-hoc comparisons showed a moderately significant advantage for VA against B, and moderately significant advantages for VS and VA against V. Furthermore, post-hoc comparisons showed a moderately marginally significant advantage for VS against B and a small yet marginally significant advantage for VA against S.

Alerting Effectiveness. Regarding alerting effectiveness, post-hoc comparisons showed substantial significant advantages for all the variants against B, substantial significant advantages for VS, VA and S against V, and moderately significant advantages for VS against VA and S.

Table 2: Results about Trust metrics across different conditions. Items: *#1. Competence; #2. Responsibility; #3. Reliability over time; #4. Degree of faith; #5. Predictability; #6. Dependability; #7. Overall trust.*

#	AVG (C.I. 95%)					*p*-value (effect size)										
	B	*V*	*S*	*VS*	*VA*	*Fried.*	*B-V*	*B-S*	*B-VS*	*B-VA*	*V-S*	*V-VS*	*V-VA*	*S-VS*	*S-VA*	*VS-VA*
1	6.13 (0.30)	6.26 (0.29)	6.26 (0.29)	6.34 (0.27)	6.21 (0.29)	.116	.233 (-0.18)	.298 (-0.18)	**.036 (-0.32)**	.571 (-0.12)	.637 (0.00)	.345 (-0.13)	.772 (0.06)	.345 (-0.13)	.772 (0.06)	.371 (0.19)
2	6.17 (0.33)	6.34 (0.21)	6.3 (0.24)	6.39 (0.21)	6.30 (0.24)	.100	.203 (-0.26)	.233 (-0.19)	.088 (-0.33)	**.049 (-0.19)**	1.00 (0.08)	1.00 (−0.08)	1.00 (0.08)	.345 (-0.16)	.637 (0.00)	1.00 (0.16)
3	6.30 (0.33)	6.30 (0.33)	6.43 (0.31)	6.47 (0.31)	6.30 (0.35)	.158	.907 (0.00)	.233 (-0.17)	.071 (-0.23)	.887 (0.00)	.371 (-0.17)	.173 (-0.23)	.841 (0.00)	1.00 (−0.05)	.148 (0.16)	.173 (0.22)
4	5.78 (0.34)	5.91 (0.25)	6.00 (0.26)	6.04 (0.27)	6.00 (0.26)	**.040**	.350 (-0.18)	.072 (-0.30)	**.041 (-0.36)**	.072 (-0.30)	.345 (-0.14)	.344 (-0.21)	.345 (-0.14)	1.00 (−0.07)	1.00 (0.00)	1.00 (0.07)
5	5.39 (0.38)	5.52 (0.36)	5.69 (0.35)	5.78 (0.36)	5.69 (0.35)	**.005**	.407 (-0.15)	.064 (-0.35)	**.017 (-0.44)**	.064 (-0.35)	.071 (-0.20)	**.047 (-0.30)**	.071 (-0.20)	1.00 (−0.10)	1.00 (0.00)	1.00 (0.10)
6	5.95 (0.24)	6.00 (0.29)	6.13 (0.23)	6.17 (0.24)	6.04 (0.27)	.096	.789 (−0.07)	.129 (-0.31)	**.036 (-0.38)**	.587 (-0.14)	.148 (-0.21)	.071 (-0.27)	.765 (-0.06)	1.00 (−0.07)	.345 (0.14)	.371 (0.21)
7	5.82 (0.3)	5.95 (0.2)	6.00 (0.22)	6.08 (0.22)	5.95 0.24)	**.051**	.298 (-0.21)	.240 (-0.27)	**.047 (-0.41)**	.482 (-0.2)	.772 (-0.08)	.148 (-0.26)	.841 (0.00)	.345 (-0.16)	.772 (0.08)	.371 (0.24)

Table 3: Result of the UEQ+ questionnaire for the critical in-ODD events (jaywalker). Items: *#1. Trustworthiness of content; #2. Quality of content; #3. Value.*

#	AVG (C.I. 95%)					*p*-value (effect size)										
	B	*V*	*S*	*VS*	*VA*	*Fried.*	*B-V*	*B-S*	*B-VS*	*B-VA*	*V-S*	*V-VS*	*V-VA*	*S-VS*	*S-VA*	*VS-VA*
1	1.22 (0.22)	1.88 (0.20)	2.05 (0.16)	2.31 (0.14)	2.15 (0.14)	**≤.001**	**≤.001 (-1.30)**	**≤.001 (-1.79)**	**≤.001 (-2.46)**	**≤.001 (-2.11)**	**.019 (-0.38)**	**≤.001 (-1.02)**	**.002 (-0.65)**	**≤.001 (-0.70)**	.104 (-0.28)	**.005 (0.45)**
2	1.24 (0.14)	1.84 (0.15)	1.89 (0.15)	2.11 (0.17)	1.99 (0.21)	**≤.001**	**≤.001 (-1.69)**	**≤.001 (-1.88)**	**≤.001 (-2.37)**	**≤.001 (-1.77)**	.354 (-0.15)	**≤.001 (-0.72)**	**.008 (-0.35)**	**≤.001 (-0.58)**	.096 (-0.23)	**.007 (0.26)**
3	1.19 (0.13)	1.67 (0.11)	1.76 (0.13)	1.92 (0.12)	1.76 (0.18)	**≤.001**	**≤.001 (-1.67)**	**≤.001 (-1.84)**	**≤.001 (-2.39)**	**≤.001 (-1.55)**	.262 (-0.29)	**.001 (-0.87)**	.089 (-0.25)	**≤.001 (-0.53)**	.643 (-0.01)	**.021 (0.43)**

Table 4: Results regarding Notification Appropriateness for the critical in-ODD events (jaywalker). Items: *#1. Message Clarity; #2. Perceived Urgency; #3. Perceived Annoyance; #4. Alerting Effectiveness.*

#	AVG (C.I. 95%)					*p*-value (effect size)										
	B	*V*	*S*	*VS*	*VA*	*Fried.*	*B-V*	*B-S*	*B-VS*	*B-VA*	*V-S*	*V-VS*	*V-VA*	*S-VS*	*S-VA*	*VS-VA*
1	2.60 (0.54)	4.86 (0.49)	6.00 (0.41)	6.78 (0.22)	5.52 (0.46)	**≤.001**	**≤.001 (-1.87)**	**≤.001 (-3.02)**	**≤.001 (-4.3)**	**≤.001 (-2.46)**	**.002 (-1.07)**	**≤.001 (-2.15)**	**.031 (-0.58)**	**≤.001 (-1.01)**	.175 (0.46)	**≤.001 (1.48)**
2	2.78 (0.88)	3.65 (0.78)	4.6 (0.62)	5.3 (0.58)	5.21 (0.59)	**≤.001**	.099 (-0.4)	**.002 (-0.97)**	**.001 (-1.29)**	**.001 (-1.21)**	**.017 (-0.57)**	**.001 (-0.92)**	**.003 (-0.85)**	**.021 (-0.40)**	.247 (-0.33)	.517 (0.05)
3	2.30 (0.7)	2.26 (0.58)	2.86 (0.71)	3.08 (0.65)	3.47 (0.79)	**.002**	.818 (0.02)	.123 (-0.34)	.055 (-0.49)	**.016 (-0.67)**	.079 (-0.40)	**.020 (-0.57)**	**.005 (-0.75)**	.255 (-0.13)	.093 (-0.34)	.209 (-0.23)
4	2.26 (0.71)	4.43 (0.64)	5.82 (0.38)	6.47 (0.31)	5.86 (0.39)	**≤.001**	**≤.001 (-1.37)**	**≤.001 (-2.68)**	**≤.001 (-3.29)**	**≤.001 (-2.69)**	**≤.001 (-1.12)**	**≤.001 (-1.72)**	**≤.001 (-1.15)**	**.001 (-0.80)**	.952 (−0.04)	**.010 (0.73)**

5.4.2 TOR Notification. Regarding TOR notification, a significant difference was found just for perceived urgency (Figure 4), whereas for the other constructs were reported non-significantly differences.

In this regard, post-hoc comparisons showed a moderately significant advantage for B against VS and small yet significant advantage for B against VA and S, consistent with the findings on anxiety perception obtained in [17], and small yet significant advantages for V against VA and VS, consistent with the findings on lower safety perception observed in [60].

Moreover, results obtained from the repeated measures two-way ANOVA analysis performed on the perceived urgency component by considering the notifications type (critical in-odd vs. TOR) and the five conditions indicated a highly significant interaction effect (*df=45*, notification type p-value < .001, conditions p-value < .001, interaction effect p-value < .001).

5.5 Ranking (Overall Preference)

The overall rank preference (Table 5) yields statistically significant differences.

Specifically, post-hoc comparisons showed substantial significant advantages for all variants against B, substantial significant advantages for VS, S and VA against V, and substantial significant

Table 5: Results of the ranking from the final questionnaire.

Rank (Overall preference)					***p*-value (effect size)**										
B	*V*	*S*	*VS*	*VA*	*Friedman*	*B-V*	*B-S*	*B-VS*	*B-VA*	*V-S*	*V-VS*	*V-VA*	*S-VS*	*S-VA*	*VS-VA*
5	4	2	1	2	**≤.001**	**.015 (0.80)**	**≤.001 (2.01)**	**≤.001 (3.25)**	**≤.001 (1.56)**	**.002 (1.25)**	**≤.001 (2.51)**	**.027 (0.90)**	**.003 (1.16)**	.816 (−0.11)	**.006 (-1.01)**

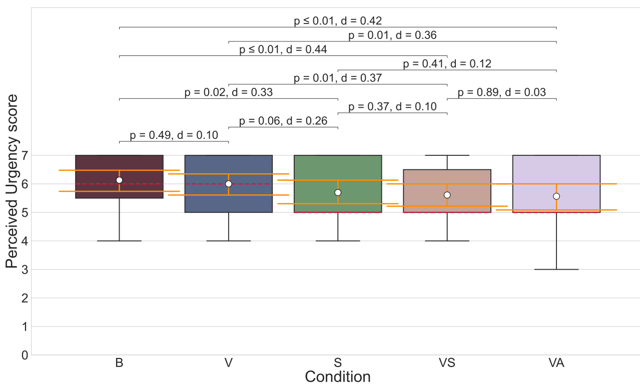


Figure 4: Results regarding TOR perceived urgency (Friedman's *p*-value<.001)

advantages for VS against S and VA, while nothing can be concluded regarding the comparison between S and VA.

5.6 Considerations & Remarks

5.6.1 Trust. The results suggest that trust levels remained consistently high across all variants of jaywalking notifications. Among the others, the VS notification demonstrated an advantage in enhancing the overall trust. The increase in trust may be attributed to the enhanced message clarity provided by the anticipation of the description of ongoing events, combined with a content framework based on "what + why". This is in line with previous studies related to information transmission [12, 17, 60, 61], stating that speech-based interfaces can significantly improve understanding of the circumstances during autonomous driving. This aspect is closely linked to the driver's sense of control over the situation and consequently contributes to the enhancement of safety and trust placed in the driving experience.

5.6.2 Critical in-ODD Events Notification. VS remained consistently high in scores as notification variant for jaywalking events in terms of message clarity, alerting effectiveness, and overall user preference while S and VA were the second-higher options. The V notification performed worse than the S notification, while B recorded the lowest values by far.

Furthermore, VS notification was also the one being perceived as more urgent together with VA, proving that the use of multimodality can improve the perceived urgency. Concerning unimodal variants, the V notification turned out to be perceived as less urgent than the S one.

Furthermore, the VA notification was perceived as the most annoying, while S and VS were perceived as slightly less annoying. These findings are consistent with previous research [5, 57], which indicated that abstract sounds increased reaction times without affecting the understanding of ongoing events. In contrast, the combination of visual feedback with the S notification can enhance message clarity and alertness effectiveness without significantly increasing the perceived annoyance of the notification.

The UEQ+ results indicate that the VS configuration was regarded as more appealing in terms of aesthetics compared to the other variants. This result may stem from the fact that speech is perceived by drivers as more anthropomorphic and, therefore, more appreciated compared to abstract and artificial sounds, in line with previous works [12]. Additionally, the visual feedback, which is based on the design previously evaluated and appreciated in an earlier study [30], complements and enriches the useful and reassuring information that a voice alone may fail to convey.

5.6.3 TOR Notifications. The results indicate that, despite the TOR notification mode was kept equal across all conditions evaluated, it has been observed that the notification mode exerted an influence on the perception of the TORs themselves. This aspect is especially reflected in changes related to the perceived urgency, highlighting how a more urgent jaywalking notification led to a less urgent perception of the TORs, albeit the magnitude of the reduction was moderate. This finding underscore the importance of considering how an event is notified, as this can influence not only the notification itself but also the perception of other notifications by individuals.

6 Conclusions

In this work, four different variants of notification approaches, two unimodal and two multimodal, were evaluated against each other and a baseline interface to investigate their appropriateness in signalling critical in-ODD events in the context of SAE L3 AV. The baseline interface was composed by an AR-WSD, a dashboard, an infotainment display, and localized sounds. A VR-based system was devised to simulate a ride in which different events occured, including the multimodal notification of TORs.

Results of the performed user study outpointed that the multimodal VS variant was overall the most preferred and able to solicit higher trust, clarity, and urgency. However, this came at the cost of a lowered perception of the TOR notification urgency. Albeit the magnitude of the detrimental effect is moderate, Future designers of such interfaces should carefully evaluate whether to compromise on variants that achieved slightly lower yet adequate notification effectiveness in order to obtain a TOR response closer to baseline, as seen in the case of the unimodal S variant.

A potential limitation may stem from the decision to have an autobraking, and eventually full AV stop, behaviour in case the TOR is not answered by the driver. Although this aspect aligns with the specifications for SAE L3 AV [44] and prior studies on TORs have already implemented this solution [11], it may diverge from the behavior that vehicles will exhibit in future real-world scenarios. Moreover, an additional limitation may arise from the selection of an NDRT that does not engage the driver's hands, such as using a mobile device. While maintaining hands-free engagement may enhance the driver's safety when regaining control, the body of literature has already raised the question of whether users, despite being warned and instructed not to do so, will still utilize devices during conditional driving.

Also testing of notifications using a VR-based simulation rather than an actual vehicle, albeit common practice in this research domain, shall be carefully considered as a limitation, as users might have perceived the simulated situations as less hazardous compared to the real world. In real-world settings with SAE L3 AV, it is speculated that magnitude of effects observed regarding trust will tend to amplify. Hence, follow-up research shall be conduced in real-world settings to further substantiate the effectiveness of the proposed multimodal notification system under uncontrolled experimental setups. It would be of particular interest, in this case, determining whether the potentially amplified perceived urgency of TORs will also translate into better reaction times and higher situation awareness. In addition, to minimize potential confounding factors, it was decided to include in the study only the jaywalker as critical in-ODD event. However, this simplification may underrepresent real-world stimuli and situations. In the future, other events of the same category shall be included in the investigation by narrowing the set of variants proposed, as to increase external validity of the results. Also increasing the variety of events and potential distractors, preferably in real-world environments, is indeed required to evaluate the system in high demand scenarios with different types of distractions and multiple source notifications, which would allow to observe the limits of driver overload and adjust the system to balance cognitive workload and safety. Such investigations could significantly contribute to the design of more effective and context-sensitive notification systems in SAE L3 AVs. Finally, further research should aim at investigating additional visual notifications, such as the use of promising vehicle interior/ambient interfaces that could complement or alternate with the already well-evaluated AR-HUDs; additionally, offered channels could be explored to complement the multimodal HMI, also considering, e.g., haptic interfaces.

Acknowledgments

This work has been carried out in the frame of the VR@POLITO initiative. The authors want to acknowledge the help provided by Mattia Mafredini in the software development.

References

[1] actuate [accessed: 02-08-2024]. Actuate Motion. https://actuate-motion.com/.

[2] Carryl L Baldwin. 2011. Verbal collision avoidance messages during simulated driving: Perceived urgency, alerting effectiveness and annoyance. *Ergonomics* 54, 4 (2011), 328–337.

[3] Paolo Bosetti, Mauro Da Lio, and Andrea Saroldi. 2014. On the human control of vehicles: An experimental study of acceleration. *European Transport Research Review* 6 (2014), 157–170.

[4] Yujia Cao, Angela Mahr, Sandro Castronovo, Mariët Theune, Christoph Stahl, and Christian A Müller. 2010. Local danger warnings for drivers: The effect of modality and level of assistance on driver reaction. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*. 239–248.

[5] Chunlei Chai, Yu Lei, Haoran Wei, Changxu Wu, Wei Zhang, Preben Hansen, Hao Fan, and Jinlei Shi. 2024. The effects of various auditory takeover requests: A simulated driving study considering the modality of non-driving-related tasks. *Applied Ergonomics* 118 (2024), 104252.

[6] Henrik Detjen, Stefan Geisler, and Stefan Schneegass. 2021. Driving as Side Task: Exploring Intuitive Input Modalities for Multitasking in Automated Vehicles. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 451, 6 pages.

[7] Cyriel Diels and Jelte E Bos. 2016. Self-driving carsickness. *Applied Ergonomics* 53 (2016), 374–382.

[8] Na Du, Jacob Haspiel, Qiaoning Zhang, Dawn Tilbury, Anuj K Pradhan, X Jessie Yang, and Lionel P Robert Jr. 2019. Look who's talking now: Implications of AV's explanations on driver's trust, AV preference, anxiety and mental workload. *Transportation Research Part C: Emerging Technologies* 104 (2019), 428–442.

[9] fanatec [accessed: 02-08-2024]. Fanatec driving kit. https://fanatec.com/eu-en/.

[10] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods* 41, 4 (2009), 1149–1160.

[11] Yannick Forster, Frederik Naujoks, and Alexandra Neukum. 2016. Your turn or my turn? Design of a human-machine interface for conditional automation. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 253–260.

[12] Yannick Forster, Frederik Naujoks, and Alexandra Neukum. 2017. Increasing anthropomorphism and trust in automated driving functions by adding speech output. In *2017 IEEE Intelligent Vehicles Symposium (IV)*. 365–372.

[13] Michael A Gerber, Ronald Schroeter, Li Xiaomeng, and Mohammed Elhenawy. 2020. Self-interruptions of non-driving related tasks in automated vehicles: Mobile vs head-up display. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–9.

[14] Thomas Alexander Goodge, Frank Pollick, and Stephen Anthony Brewster. 2024. Can You Hazard a Guess?: Evaluating the Effect of Augmented Reality Cues on Driver Hazard Prediction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–28.

[15] Renate Haeuslschmid, Yixin Shou, John O'Donovan, Gary Burnett, and Andreas Butz. 2016. First steps towards a view management concept for large-sized head-up displays with continuous depth. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 1–8.

[16] hmitemp [accessed: 02-08-2024]. Unity Automotive HMI Template. https://docs.unity3d.com/templates/automotive-hmi-template@1.0/.

[17] Ann Huang, Shadi Derakhshan, John Madrid-Carvajal, Farbod Nosrat Nezami, Maximilian Alexander Wächter, Gordon Pipa, and Peter König. 2024. Enhancing Safety in Autonomous Vehicles: The Impact of Auditory and Visual Warning Signals on Driver Behavior and Situational Awareness. *Vehicles* 6, 3 (2024), 1613–1636.

[18] Quinate Chioma Ihemedu-Steinke, Prashanth Halady, Gerrit Meixner, and Michael Weber. 2018. VR evaluation of motion sickness solution in automated driving. In *Virtual, Augmented and Mixed Reality: Interaction, Navigation, Visualization, Embodiment, and Simulation: 10th International Conference, VAMR 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I 10*. Springer, 112–125.

[19] Julie Iskander, Mohammed Attia, Khaled Saleh, Darius Nahavandi, Ahmed Abobakr, Shady Mohamed, Houshyar Asadi, Abbas Khosravi, Chee Peng Lim, and Mohammed Hossny. 2019. From car sickness to autonomous car sickness: A review. *Transportation Research Part F: Traffic Psychology and Behaviour* 62 (2019), 716–726.

[20] Christian P Janssen, Shamsi T Iqbal, Andrew L Kun, and Stella F Donker. 2019. Interrupted by my car? Implications of interruption and interleaving research for automated vehicles. *International Journal of Human-Computer Studies* 130 (2019), 221–233.

[21] Christian P Janssen, J Leon Kenemans, et al. 2015. Multitasking in autonomous vehicles: Ready to go?. In *3rd Workshop on User Experience of Autonomous Vehicles at AutoUI'*. ACM Press.

[22] Gunnar Johansson and Kåre Rumar. 1971. Drivers' brake reaction times. *Human Factors* 13, 1 (1971), 23–27.

[23] Richie Jose, Gun A Lee, and Mark Billinghurst. 2016. A comparative study of simulated augmented reality displays for vehicle navigation. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction*. 40–48.

[24] Nihan Karatas, Takahiro Tanaka, Kazuhiro Fujikakc, Yuki Yoshihara, Hitoshi Kanamori, Yoshitaka Fuwamoto, and Morihiko Yoshida. 2020. Evaluation of AR-HUD interface during an automated intervention in manual driving. In *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2158–2164.

[25] Soyeon Kim, René van Egmond, and Riender Happee. 2024. How manoeuvre information via auditory (spatial and beep) and visual UI can enhance trust and acceptance in automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour* 100 (2024), 22–36.
[26] Thomas Köhn, Matthias Gottlieb, Michael Schermann, and Helmut Krcmar. 2019. Improving take-over quality in automated driving by interrupting non-driving tasks. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 510–517.
[27] Sanguk Lee, Rabindra Ratan, and Taiwoo Park. 2019. The voice makes the car: Enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style. *Multimodal Technologies and Interaction* 3, 1 (2019), 20.
[28] Bridget A Lewis, Jesse L Eisert, and Carryl L Baldwin. 2018. Validation of essential acoustic parameters for highly urgent in-vehicle collision warnings. *Human factors* 60, 2 (2018), 248–261.
[29] Shuo Li, Phil Blythe, Simon Edwards, Paul Goodman, and Graeme Hill. 2019. Investigation of the influence of multitasking on drivers' takeover performance in highly automated vehicles. In *26th Intelligent Transport Systems World Congress*. Newcastle University.
[30] Patrick Lindemann, Tae-Young Lee, and Gerhard Rigoll. 2018. Supporting driver situation awareness for autonomous urban driving with an augmented-reality windshield display. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. 358–363.
[31] Lia Morra, Fabrizio Lamberti, F Gabriele Pratticó, Salvatore La Rosa, and Paolo Montuschi. 2019. Building trust in autonomous vehicles: Role of virtual reality driving simulators in HMI design. *IEEE Transactions on Vehicular Technology* 68, 10 (2019), 9438–9450.
[32] Clifford Nass, Ing-Marie Jonsson, Helen Harris, Ben Reaves, Jack Endo, Scott Brave, and Leila Takayama. 2005. Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. 1973–1976.
[33] Frederik Naujoks, Katharina Wiedemann, and Nadja Schömig. 2017. The importance of interruption management for usefulness and acceptance of automated driving. In *Proceedings of the 9th International Conference on automotive User Interfaces and Interactive Vehicular Applications*. 254–263.
[34] Teo Niemirepo, Juuso Toivonen, Marko Viitanen, and Jarno Vanne. 2019. Opensource CiThruS simulation environment for real-time 360-degree traffic imaging. In *2019 IEEE International Conference on Connected Vehicles and Expo (ICCVE)*. 1–5.
[35] Luis Oliveira, Jacob Luton, Sumeet Iyer, Chris Burns, Alexandros Mouzakitis, Paul Jennings, and Stewart Birrell. 2018. Evaluating how interfaces influence the user interaction with fully autonomous vehicles. In *Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 320–331.
[36] openmc [accessed: 02-08-2024]. OpenVR Motion Compensation. https://ovrmc.dschadu.de/.
[37] Michalis Papakostas, Kais Riani, Andrew Brian Gasiorowski, Yan Sun, Mohamed Abouelenien, Rada Mihalcea, and Mihai Burzo. 2021. Understanding driving distractions: A multimodal analysis on distraction characterization. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 377–386.
[38] Sebastiaan Petermeijer, Pavlo Bazilinskyy, Klaus Bengler, and Joost De Winter. 2017. Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics* 62 (2017), 204–215.
[39] Ioannis Politis, Stephen A Brewster, and Frank Pollick. 2014. Evaluating multimodal driver displays under varying situational urgency. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 4067–4076.
[40] Edita Poljac, Andrea Kiesel, Iring Koch, and Hermann Müller. 2018. New perspectives on human multitasking. *Psychological Research* 82, 1 (2018), 1–3.
[41] Jonathan J Rolison, Shirley Regev, Salissou Moutari, and Aidan Feeney. 2018. What are the factors that contribute to road accidents? An assessment of law enforcement views, ordinary drivers' opinions, and road accident records. *Accident Analysis & Prevention* 115 (2018), 11–24.
[42] Benjamin S. Alpers, Kali Cornn, Lauren E. Feitzinger, Usman Khaliq, So Yeon Park, Bardia Beigi, Daniel Joseph Hills-Bunnell, Trevor Hyman, Kaustubh Deshpande, Rieko Yajima, et al. 2020. Capturing passenger experience in a ride-sharing autonomous vehicle: The role of digital assistants in user interface design. In *12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. 83–93.
[43] Shadan Sadeghian Borojeni, Susanne CJ Boll, Wilko Heuten, Heinrich H Bülthoff, and Lewis Chuang. 2018. Feel the movement: Real motion influences responses to take-over requests in highly automated vehicles. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.
[44] Spencer Salter, Cyriel Diels, Paul Herriotts, Stratis Kanarachos, and Doug Thake. 2019. Motion sickness in automated vehicles with forward and rearward facing seating orientations. *Applied Ergonomics* 78 (2019), 54–61.
[45] Kevin Joel Salubre and Dan Nathan-Roberts. 2021. Takeover request design in automated driving: A systematic review. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 65. SAGE Publications Sage CA: Los Angeles, CA, 868–872.
[46] Dario D Salvucci, Niels A Taatgen, and Jelmer P Borst. 2009. Toward a unified theory of the multitasking continuum: From concurrent performance to task switching, interruption, and resumption. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1819–1828.
[47] Tobias Schneider, Sabiha Ghellal, Steve Love, and Ansgar RS Gerlicher. 2021. Increasing the user experience in autonomous driving through different feedback modalities. In *Proceedings of the 26th International Conference on Intelligent User Interfaces*. 7–10.
[48] Martin Schrepp, Heike Sandkühler, and Jörg Thomaschewski. 2021. How to create short forms of UEQ+ based questionnaires? (2021).
[49] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. 2017. AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles. In *Field and Service Robotics*. 621–635.
[50] Elisabeth Shi and Alexander T Frey. 2021. Non-Driving-Related Tasks During Level 3 Automated Driving Phases—Measuring What Users Will Be Likely to Do. *Technology, Mind, and Behavior* (2021).
[51] Lee Skrypchuk, Pat Langdon, Ben D. Sawyer, Alex Mouzakitis, and P. John Clarkson. 2019. Enabling multitasking by designing for situation awareness within the vehicle environment. *Theoretical Issues in Ergonomics Science* 20, 2 (2019), 105–128.
[52] Pinyan Tang, Xu Sun, and Shi Cao. 2020. Investigating user activities and the corresponding requirements for information and functions in autonomous vehicles of the future. *International Journal of Industrial Ergonomics* 80 (2020), 103044.
[53] tracker [accessed: 02-08-2024]. Vive Tracker 3.0. https://www.vive.com/nz/accessory/vive-tracker/.
[54] unity [accessed: 02-08-2024]. Unity. https://unity.com.
[55] vive [accessed: 02-08-2024]. HTC VIVE Pro kit. https://www.vive.com/us/product/vive-pro-full-kit/.
[56] waymo [accessed: 02-08-2024]. Waymo robotaxi service in Los Angeles. https://waymo.com/waymo-one-los-angeles/.
[57] Timothy J Wright, Ravi Agrawal, Siby Samuel, Yuhua Wang, Shlomo Zilberstein, and Donald L Fisher. 2017. Effects of alert cue specificity on situation awareness in transfer of control in level 3 automation. *Transportation Research Record* 2663, 1 (2017), 27–33.
[58] Weiyin Xie, Dengbo He, and Genhao Wu. 2023. Inducers of motion sickness in vehicles: A systematic review of experimental evidence and meta-analysis. *Transportation Research Part F: Traffic Psychology and Behaviour* 99 (2023), 167–188.
[59] Yucheng Yang, Burak Karakaya, Giancarlo Caccia Dominioni, Kyosuke Kawabe, and Klaus Bengler. 2018. An HMI concept to improve driver's visual behavior and situation awareness in automated vehicle. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 650–655.
[60] Sol Hee Yoon, Young Woo Kim, and Yong Gu Ji. 2019. The effects of takeover request modalities on highly automated car control transitions. *Accident Analysis & Prevention* 123 (2019), 150–158.
[61] Qiaoning Zhang, X Jessie Yang, and Lionel P Robert. 2021. What and when to explain? A survey of the impact of explanation on attitudes toward adopting automated vehicles. *IEEE Access* 9 (2021), 159533–159540.
[62] Yimin Zhou and Meng Xu. 2023. Robotaxi service: The transition and governance investigation in China. *Research in Transportation Economics* 100 (2023), 101326.