



Article

Displays for Productive Non-Driving Related Tasks: Visual Behavior and Its Impact in Conditionally Automated Driving

Clemens Schartmüller ^{1,2,*} , Klemens Weigl ^{1,3} , Andreas Löcken ¹ , Philipp Wintersberger ^{1,4} ,
Marco Steinhauser ³ and Andreas Riener ^{1,2}

¹ CARISSMA Institute of Automated Driving, Technische Hochschule Ingolstadt (THI), 85049 Ingolstadt, Germany; klemens.weigl@thi.de (K.W.); andreas.loecken@carissma.eu (A.L.); andreas.riener@thi.de (A.R.)

² Institute for Pervasive Computing, Johannes Kepler University Linz (JKU), 4020 Linz, Austria

³ Department of Psychology, Catholic University of Eichstätt-Ingolstadt, 85072 Eichstätt, Germany; marco.steinhauser@ku.de

⁴ Institute of Visual Computing and Human-Centered Technology, TU Wien, 1040 Vienna, Austria; philipp.wintersberger@tuwien.ac.at

* Correspondence: clemens.schartmueller@thi.de

Abstract: (1) Background: Primary driving tasks are increasingly being handled by vehicle automation so that support for non-driving related tasks (NDRTs) is becoming more and more important. In SAE L3 automation, vehicles can require the driver-passenger to take over driving controls, though. Interfaces for NDRTs must therefore guarantee safe operation and should also support productive work. (2) Method: We conducted a within-subjects driving simulator study ($N = 53$) comparing Heads-Up Displays (HUDs) and Auditory Speech Displays (ASDs) for productive NDRT engagement. In this article, we assess the NDRT displays' effectiveness by evaluating eye-tracking measures and setting them into relation to workload measures, self-ratings, and NDRT/take-over performance. (3) Results: Our data highlights substantially higher gaze dispersion but more extensive glances on the road center in the auditory condition than the HUD condition during automated driving. We further observed potentially safety-critical glance deviations from the road during take-overs after a HUD was used. These differences are reflected in self-ratings, workload indicators and take-over reaction times, but not in driving performance. (4) Conclusion: NDRT interfaces can influence visual attention even beyond their usage during automated driving. In particular, the HUD has resulted in safety-critical glances during manual driving after take-overs. We found this impacted workload and productivity but not driving performance.

Keywords: automated driving; take-over requests; conditional automation; non-driving related tasks; eye-tracking; performance; behavior; displays; productivity; visual attention



Citation: Schartmüller, C.; Weigl, K.; Löcken, A.; Wintersberger, P.; Steinhauser, M.; Riener, A. Displays for Productive Non-Driving Related Tasks: Visual Behavior and Its Impact in Conditionally Automated Driving. *Multimodal Technol. Interact.* **2021**, *5*, 21. <https://doi.org/10.3390/mti5040021>

Academic Editor: Mu-Chun Su

Received: 15 February 2021

Accepted: 13 April 2021

Published: 18 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

After typical driver assistance systems (SAE L1, e.g., adaptive cruise control; cf., [1]) and combinations thereof (SAE L2, partial automation), the upcoming level of driving automation is called *conditional automation* (SAE L3) and will no longer require drivers to monitor the driving system at all times. Thus, driver distraction by so-called Non-Driving Related Tasks (NDRTs) becomes less of a concern. However, the *driver-passenger* is still required to respond to the so-called Take-Over Requests (TORs) issued by the SAE L3 vehicle in cases where the system exceeds its limits [1] and hence re-engage in manual vehicle control. Research reports a detrimental effect of NDRT engagement on the driver's performance regarding, for example, reaction times or steering behavior (cf., [2–4]). This impact is especially detrimental with visual distraction [5]. In contrast to lower levels of automation (SAE L0–L2), reports are not so clear-cut. They suggest that NDRTs could even be required to prevent cognitive underload, drowsiness, mind-wandering, or effects of

boredom (cf., [6–9]). Recent research further shows the desirability of NDRTs in automated vehicles (cf., [10,11]), and people will likely engage in NDRTs, even if prevented by legislation, as can already be observed in today's driving below L3 automation [12]. Hence, we believe that laws preventing NDRTs are not the answer to potential safety issues when NDRTs are promoted from a secondary to the second primary task (besides maintaining take-over readiness), as soon as high levels of automation reach the market (SAE L3 to L5). Instead, we and others [13,14] argue that well-integrated and optimized NDRT user interfaces (NDRT-UIs) are needed that optimize cognitive, perceptual, and motor processes to enable safe, productive, and enjoyable NDRT engagement in SAE L3 vehicles. Especially demanding NDRTs that go beyond previous limits of the occasional short text message or switching the radio station, such as when using the time on the road to work [15], will require special attention by designers and researchers.

Strategies for NDRT-UI optimization include managing visual attention with various display locations [16], behaviors [17], and modalities [18]. It is necessary to assess the effects of these variants on driver-vehicle interaction first to design UIs in a safe, productive, and enjoyable way.

In this work, we investigate how two different display variants, namely Heads-Up Displays (HUDs) and Auditory Speech Displays (ASDs), affect visual attention and behavior in automated driving followed by critical TOR situations. Additionally, we assessed the potential for productivity in an office-like text comprehension task, safety during take-overs, as well as overall enjoyability of the display variants.

1.1. In-Vehicle Displays

Displays in vehicles typically serve either the purpose of providing driving-related (such speed indicators or system warnings) or non-driving related information (such as media or call applications).

Typical are analog or digital instrument clusters behind the steering wheel, in-vehicle infotainment systems in the center console, auditory notifications, including limited spoken information (e.g., navigational instructions or short messages text-to-speech), and sometimes haptic information.

More recently, production vehicles include Heads-Up Displays that provide driving-related information near to, or overlaying, the road on the windshield, which are found to mitigate issues of cognitive load and visual attention [18]. Upcoming generations of such HUDs could allow sophisticated augmented reality applications, for example, to foster user trust [19,20] but might also be useful for a wide range of NDRTs [21]. Still, HUDs positioned at fixed depths, may introduce a visual accommodation delay (eye focus) and even lead to missing, i.e., not perceiving objects within the user's field of view [22], which may diminish their potential benefits with spatial proximity of information, in comparison to traditional visual displays such as the instrument cluster or center stack infotainment display.

Auditory feedback systems are continuously extended in their capabilities too; for example, as display modalities for interactive digital assistants. For text-based NDRTs and navigational instructions, Auditory Speech Displays have been implemented with text-to-speech systems in cars, as they do not require visual attention for, e.g., reading out short text messages. Research also explores the sonification of extended information beyond text messages [23] using concepts such as spearcons [24] and lyricons [25].

Haptic feedback is prominently used for warnings and driving guidance [26] in the form of vibrotactile cues since it can cause annoyance [27] and is limited in its information density. Transfer of extended amounts of information, such as when displaying text for NDRTs, would require most people to learn reading Braille or similar tactile text systems (e.g., in the United Kingdom only about 20,000 people are estimated to read Braille [28]). We thus do not investigate haptics further in this study.

Overall, Large et al. [29] and others argue that retaining the required situation awareness for lower levels of automation may suffer specifically from visual distraction. In a similar fashion, for example, You et al. [30] have shown that a voice chat can more positively

influence users' ability to handle take-overs than visual reading. Chang et al. [31] show that active speaking is still causing significantly more workload than listening, though.

In summary, neither of these studies did compare HUDs and ASDs directly, but rather always separately to a "heads-down" baseline, e.g., on the center stack or on a handheld smartphone. So the question arises, which variant is more suitable for productivity-oriented NDRTs in conditionally automated vehicles, and where are the differences/how could they be improved.

1.2. Theoretical Considerations

The driver's challenge in a SAE L3 vehicle lies in performing well in an NDRT, such as reading, and, in parallel, remain ready to respond well to a take-over request. The Multiple Resource Theory by Wickens [32] suggests that an auditory display, such as an ASD, is favorable over a visual display since the NDRT is shifted away from the visual channel, which is already occupied by observing the driving environment, to the auditory channel. At the same time, the Cognitive Control Hypothesis by Engström et al. [33] would suggest that performance may not be undermined if the task is well-practiced and consistently mapped, as could be the case with a properly adapted visual display, such as a HUD. Also, people are well accommodated to visual interaction (especially for office tasks) and therefore might perform better with this as NDRT display modality. However, the cost of performing the task visually is a potential loss of situation awareness (SA). SA can be described as the degree to which an operator is aware of a process. Endsley [34] distinguishes three levels—the perception of elements relevant for the situation (level 1), the comprehension of this situation (level 2), and the ability to predict future states (level 3). Visual behavior/eye-tracking has been used in various works to infer situation awareness (cf., [35–37]). Although it cannot be guaranteed that an operator has comprehended and can predict future states of a situation just by looking at it, the analysis of eye-tracking measures allows investigating if a situation is perceived (i.e., SA level 1), an important contributor to establishing sufficient SA.

Summarizing theory, we would expect the ASD to have significant advantages in terms of workload distribution and thereby performance, as suggested by Wickens's Multiple Resource Theory. However, Engström et al.'s Cognitive Control Hypothesis suggests that this effect can be alleviated if the interfaces are adapted well enough, leading us to ask whether HUDs are better adapted visual displays, in comparison to ASDs. Concretely, performance in our setting would be characterized by safety (take-over reaction and driving performance), productivity (NDRT performance), and enjoyability (user ratings of acceptance and workload). A potential cause for the display variants' effects on these aspects could be highlighted by investigating the compared display variants' most distinctive requirement: visual attention.

1.3. Research Questions and Hypotheses

Both HUDs and ASDs have already been employed in vehicles on the road for driving and non driving-related information, but systems supporting more demanding NDRTs, such as office work tasks, have not yet been realized and deployed "in the wild". To assess the applicability of these two display variants for more demanding NDRTs—in particular, reading as part of office work—we compared them in a lab study. We already discussed the impact on productivity, stress, workload, take-over performance, and user acceptance in [38] for the first 32 participants. In this first evaluation, we found that the HUD provided more productivity than the ASD, whereas the ASD was subjectively favored, possibly because gaze was less restricted with the ASD during automated driving. To further investigate the impact of these NDRT display modalities, we extended the sample and performed more detailed analyses on visual behavior. This work, therefore, focuses on the impact of using the two NDRT display variants (HUD or ASD) on users' visual behavior during automated driving with NDRT and also during subsequent take-overs. We recruited 21 more participants and thus present the analysis on the complete set of $N = 53$

participants, focusing on visual behavior. Then, we discuss our results as a potential reason for the previously published findings, to infer implications for human-machine interaction in conditionally automated vehicles.

We propose two research questions and present a set of hypotheses for investigation:

RQ1 How do the NDRT display variants (HUD and ASD) influence drivers' glance behavior?

RQ2 What can we learn from combining insights of our investigation of physiological workload, TOR and NDRT performance, and self-ratings, with visual behavior, to foster safe, productive and enjoyable human-computer interaction in conditionally automated vehicles?

We postulate that the choice of the display (HUD vs. ASD) influences drivers' monitoring behavior and ultimately their response to a TOR:

Hypothesis 1 (H1). *The HUD requires visual attention for the NDRT, which leads to visual attention focused on the center portion of the windshield, divided between the road center and HUD. In turn, with the ASD, drivers spend more visual attention to other driving-related information areas during automated driving.*

Hypothesis 2 (H2). *With the ASD, drivers spend more visual attention to areas that are not related to either of the two primary tasks (maintaining situation awareness and the NDRT) during automated driving than with HUD.*

Hypothesis 3 (H3). *Preceding ASD usage requires less visual scanning of the environment to regain sufficient situation awareness upon a TOR, compared to the HUD.*

Hence, we analyze visual behavior thoroughly in this article and discuss it jointly with the updated analysis of our previous findings. Furthermore, we analyzed the impact of user characteristics and pre-study experiences, but this is beyond the scope of this article and will be subject to a future publication.

2. Materials and Methods

We conducted a driving simulator user study with a total of $N = 53$ participants (extended from $N = 32$ in [38]), using a within-subjects design with the two NDRT display variants (HUD vs. ASD) as the independent variable. It was presented to participants on a *drive-level*—i.e., each drive used a different display variant in a randomized manner. The procedure is depicted in Figure 1 and described in the following.

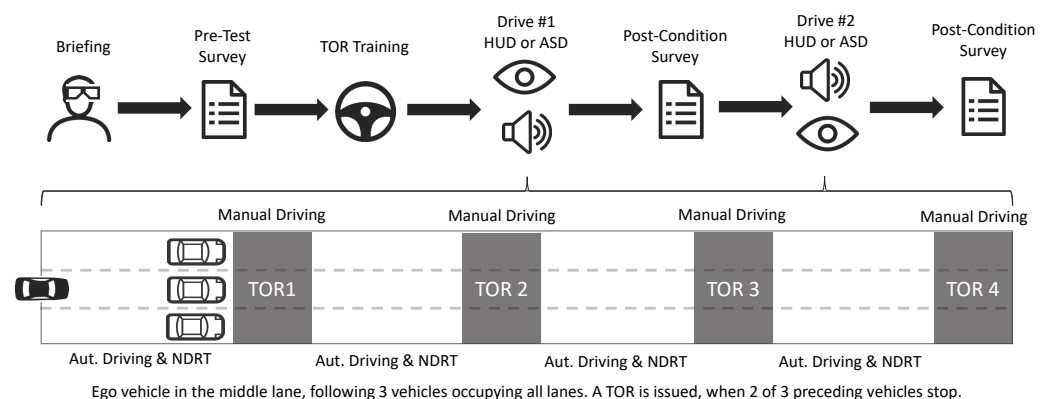


Figure 1. Driving simulator study procedure involving the two randomized conditions HUD and ASD (top), with 4 automated driving and 4 take-over phases per condition (bottom).

2.1. Implementation

As a basis for the study's implementation, we used a hexapod driving simulator with an instrumented VW Golf cockpit. U.S.E. [39] as lab software allowed prototypes, driving

simulation, and physiological measurements to be synchronized and further, to automate each study run execution, such that no experimenter intervention was necessary during the simulated drives. The in-cockpit setup is shown and described in Figure 2.



Figure 2. Driving Simulator Cockpit from participants' perspective (wide-angle view), with (1) GSR electrodes attached to the left hand's middle and ring finger, (2) observation camera, (3) driving mode display and (4) the computer mouse for task-responses. The HUD, which was only visible in the HUD condition, is highlighted with a red border.

Prototype

A semi-transparent (50% opacity) white overlay, centered in the upper third and left two thirds of the windshield served as HUD (cf., shown in Figure 2). The positioning was chosen as per recommendations of [40] in the top section overlaying the sky. This positioning also allows analyzing visual multitasking behavior as the HUD does not overlay the road.

As ASD, we used the freely available Google Text-To-Speech Engine with default female voice. Here, participants could choose between 7 different speed levels ($0.5 \times - 3.0 \times$ the default speed) prior to the drive with the ASD.

Upon a TOR, in condition ASD the auditory NDRT output was stopped and the HUD disappeared respectively in condition HUD. Sentences were interrupted in half of all TOR situations and consequently discarded (i.e., also excluded from NDRT performance recordings). The NDRT was resumed with a new sentence after the TOR situation was successfully handled and automated driving resumed (cf., Section 2.2.1).

2.2. Procedure

As illustrated in Figure 1, participants in the beginning received, read, and signed a consent form and written introduction to the experiment and were then equipped with the physiological measurement devices. In the driving simulator, they underwent a short training session with two take-overs, accompanied by the experimenter explaining the procedure. Subsequently, the two test-drives followed, each with the respective condition (HUD or ASD) in quasi-randomized order. Each drive then consisted of four scenarios, including two driving phases each: 1. automated driving (AD) with NDRT engagement → 2. TOR followed by manual driving (MD) that is ended by a successful handback to AD). During AD, participants had to perform the NDRT. After each drive, participants had to fill a range of self-rating scales on workload and acceptance on a tablet computer directly in the simulator cockpit to reflect on their experiences during the drive. Participants were debriefed after both conditions and received a compensation of 20 Euros.

2.2.1. Take-Over Task

Participants drove on a three-lane German layout highway at 120 km/h with three vehicles in front of them (illustrated in Figure 1), and always started a drive in AD mode on the middle lane. The current driving mode and take-over requests were displayed to the user on a 10 inch tablet PC in the center stack (i.e., the driving mode display, DMD; cf., Figure 2). In case of a TOR, the DMD visually changed to “Please Take-Over” shown on red background and accompanied by an auditory “beep”. Additionally, the respective NDRT display was hidden (HUD) or stopped (ASD). We deliberately chose to separate the TOR from the NDRT display (i.e., not visually display the TOR on the HUD or with spoken text in the ASD condition), to be able to clearly identify the NDRT display’s impact, as such, on the user.

A TOR was issued when two of the three vehicles in front (i.e., the middle, and either left or right one) stopped. All TORs occurred at a time-to-collision of 5 s, representing a so-called urgent take-over request due to its emergency nature [17]. After taking over control (by actuating a pedal for >5 % or a steering angle change of >2 °, cf. [41]), participants had to manually drive from the middle lane to the free lane (i.e., either the very left or right lane) to avoid a collision. To hand back control to AD, the car had to be aligned in the middle lane for 2.5 s without interruption. During this maneuver, the DMD showed “Manual Driving” on gray background with a yellow progress bar filling up, when the car was correctly aligned in the middle lane. The timer and progress bar reset when this maneuver was interrupted by leaving the middle lane. Successfully completing the handback maneuver caused the vehicle to resume AD.

2.2.2. Text Comprehension Task

The NDRT during AD consisted of rating the semantic correctness of single sentences on a binary scale, displayed to participants via the respective condition’s NDRT display modality. The task is based on the reading-span task by Daneman and Carpenter [42], and was chosen because it shares similarities with typical office tasks, such as proofreading, and is still comparable between conditions and users. It included about 80 sentences per condition, which participants were instructed to rate as fast but also as correct as possible on their semantic correctness. A rating was issued for each sentence via a button press on a computer mouse in the center console (cf., Figure 2; left button highlighted in green: “semantically correct”, right button in red: “semantically incorrect”) within 10 s. For each sentence-response or a timeout, feedback was given via the respective NDRT display modality (green check mark/“correct”, red cross/“incorrect”, cartoon-snail/“too slow”). If a TOR interrupted a sentence, then the sentence was discarded.

2.3. Visual Behavior Measurements

Tobii Pro Glasses 2, recording at 100 Hz and corrective lenses (cf., [43]), served as head-mounted eye-tracker with iMotions as software-basis (cf., [44]) for gaze and fixation analysis. The eye-tracker was calibrated before each condition.

Analyzing visual behavior was noted to be crucial to ensure safe and successful transitions out of automated driving by presenting insights on multitasking behavior [45]. Following ISO 15007 [46], a glance is defined by “maintaining of visual gaze within an area of interest (AoI), bounded by the perimeter of the area of interest”, whereas a single glance typically spans the time from the first fixation at the AoI until the first fixation away from it. All AoIs analyzed in this work are illustrated in Figure 3.



Figure 3. Analyzed areas of interest (AoIs). Also introduced was an “Other”-AoI as target of all fixations which were not within the bounding boxes of the discretely defined AoIs.

To answer our research questions and validate our hypotheses, we investigated the following behavioral measures per participant and condition (ASD vs. HUD; driving phases automated driving and take-overs were analyzed separately). Specific to single AoIs are:

Mean Glance Duration: mean glance duration in milliseconds.

Glance Rate: glances towards an AoI per second, also known as glance frequency.

Glance Location Probability: number of glances to/in the respective AoI, relative to the total number of glances.

To analyze visual attention dispersion independently from AoIs, we calculated the **Horizontal and Vertical Gaze Dispersion**, i.e., the absolute deviation of gaze from the very road center in degrees. It was linked to driver distraction and workload [47]. Further included in the analysis are **Link Value Probabilities** from/to the RoadCenter. These assess the probability of a transition from/to the road center in relation to the number of all recorded transitions [46].

General visual distraction measures for both driving phases included are based on ISO 15007 [46], Victor et al. [48], and Green et al. [49]:

Percentage Road Center (PRC): total glance time on the AoI RoadCenter (relative to the driving phase duration).

Percentage Eyes Off Road Time (PEORT): total glance time outside of AoIs Road and RoadCenter (relative to the driving phase duration).

Percentage Transition Time: time spent during which eyes are in movement between AoIs (relative to the driving phase duration).

Percentage Extended Glances Off Road (PEGOR): percentage of extended (>2 s long) glances outside the AoIs Road and RoadCenter. Evaluated during take-overs only.

To assess H1, we evaluate glancing and glance transitioning, gaze dispersion, and visual distraction measures concerning the road center and displays with driving-related

information that can be used to anticipate an upcoming TOR during automated driving, i.e., AoIs Road Center, InstrumentCluster, and DMD (cf. Figure 3).

To assess H2, we evaluate glancing and glance transition measures concerning glances to areas that are supposedly not relevant to either of the two primary tasks (maintaining situation awareness and the NDRT), i.e., AoI “Other”, during automated driving.

To assess H3, we evaluate visual distraction measures in the take-over phase, i.e., from the point in time a TOR is issued until manual driving is successfully accomplished by completing the handback maneuver.

2.4. Measurements for Safety, Productivity, Workload, and Acceptance

To assess safety, we recorded TOR reaction times and driving performance measures [50]. For productivity, we evaluated users’ NDRT performance [51]. Workload indicators used were self-ratings (raw NASA TLX, [52]) and two physiological recordings (Skin Conductance Level, Pupil Diameter, [53]). Acceptance of the investigated technologies was assessed with self-ratings based on the Technology Acceptance Model (TAM, [54]).

Self-rating measures were recorded after each drive. Physiological and performance measures were recorded as average of the four driving phases of the same type (i.e., automated driving with NDRT engagement and take-overs with manual driving from the issuing of a TOR until the successful handback maneuver) in each drive. For a detailed description of these specific measures and their calculation procedures, we refer to [38].

2.5. Software and Tests for Statistical Analyses

For statistical analyses, IBM SPSS 25 [55] was used. In most cases, variance homogeneity, and normality was met. Additionally, we assessed skewness and kurtosis and examined the non-normal data graphically with QQ-plots, which revealed satisfactory results. Moreover, the central limit theorem ($N \geq 30$) applies for our sample ($N = 53$) and relaxes the rather strict assumption of normality, which is more critical for smaller samples. Hence, we applied parametric statistical analyses. For multiple comparisons, we adjusted the significance level from $\alpha = 0.05$ to the Bonferroni-corrected α' . For multi-item self-rating dimensions, we computed Cronbach’s Alpha (CA), resulting in satisfying reliability scores for all “Raw” NASA TLX and TAM dimensions. Abbreviations M for mean, Mdn for median, SD for standard deviation, and SE for standard error are used throughout the article.

2.6. Participants

We recruited 53 participants, 27 female and 26 male ($M_{age} = 29.29$, $SD_{age} = 12.91$), via local newspaper advertisements and mailing lists. All participants owned a valid driving license (self-estimated mileage per year: $M = 12,781.25$ km, $SD = 9816.82$ km), were native or fluent in the NDRT’s language (German), reported no psychiatric disorder, and to be under no alcohol or drug influence. Furthermore, they also had normal or corrected-to-normal vision (using the eye-tracker’s corrective lenses kit) and no prior experience with take-overs in SAE Level 3 driving.

3. Results

This section summarizes the findings of statistical hypotheses testing with the full sample ($N = 53$). The discussion of these results in relation to the research questions, hypotheses, and related work, is in Section 4.

Descriptive and test statistics with effect sizes are listed in Table 1 summarizing automated driving phases and Table 2 for take-over phases in the respective conditions. Considering effect sizes, we refer to the classification of the effect size of Cohen’s d [56], which states values of approximately 0.2, 0.5, or 0.8 refer to a small, medium, and large effect, respectively. Therefore, all significant p values in Tables 1 and 2 indicate at least a medium or a large effect. The effect sizes of above 1 refer to a sufficiently large effect and a mean difference of more than one standard deviation.

Table 1. Descriptive statistics of visual behavior measures during *automated driving* in the conditions heads-up display (HUD) and auditory speech display (ASD). We report the mean (M), standard deviation (SD), *t*-test statistics, and effect sizes (Cohen's *d*). AoI driving mode display (DMD) was excluded from the AoI-specific analysis due to the low number of samples observed.

AUTOMATED DRIVING							
		HUD		ASD			
	Variable	M (SD)	M (SD)	df	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Glances at RoadCenter	Mean Glance Dur. (ms)	2240.67 (1996.95)	12,323.02 (16205.70)	47	−4.32	0.000 *	0.620
	Glance Rate (per sec.)	0.17 (0.05)	0.11 (0.05)	47	7.97	0.000 *	1.15
	Glance Loc. Prob. (rel.)	0.38 (0.08)	0.49 (0.08)	47	−7.90	0.000 *	1.14
Glances at HUD	Mean Glance Dur. (ms)	1861.62 (561.13)	-	-	-	-	-
	Glance Rate (per sec.)	0.21 (0.06)	-	-	-	-	-
	Glance Loc. Prob. (rel.)	0.43 (0.06)	-	-	-	-	-
Glances at InstrumentCluster	Mean Glance Dur. (ms)	940.79 (486.26)	2498.92 (7075.92)	32	−1.24	0.224	0.22
	Glance Rate (per sec.)	0.02 (0.04)	0.02 (0.02)	47	−0.04	0.971	0.01
	Glance Loc. Prob. (rel.)	0.03 (0.08)	0.07 (0.08)	47	−2.73	0.009 *	0.76
Glances at Other	Mean Glance Dur. (ms)	904.90 (850.47)	1751.25 (4376.21)	47	−1.29	0.202	0.19
	Glance Rate (per sec.)	0.08 (0.04)	0.07 (0.04)	47	0.55	0.588	0.08
	Glance Loc. Prob. (rel.)	0.16 (0.07)	0.28 (0.10)	47	−7.01	0.000 *	1.01
Visual Distraction	Perc. Transition Time (rel.)	0.10 (0.05)	0.09 (0.07)	47	0.71	0.479	0.10
	PEORT	0.62 (0.16)	0.30 (0.17)	47	10.84	0.000 *	1.51
	PRC	0.36 (0.14)	0.64 (0.18)	47	−9.15	0.000 *	1.32
Link Value Prob. from/to RoadCenter	to/from Other	0.07 (0.04)	0.49 (0.22)	47	−13.43	0.000 *	1.94
	to/from Road	0.06 (0.15)	0.29 (0.15)	47	−9.56	0.000 *	1.38
	to/from TOR	0.0003 (0.001)	0.0038 (0.009)	47	−2.69	0.010 *	0.39
	to/from InstrumentCluster	0.01 (0.02)	0.05 (0.05)	47	−6.41	0.000 *	0.92
Gaze Dispersion	Horizontal (degr.)	8.74 (2.78)	11.91 (3.14)	47	−5.978	0.000 *	0.864
	Vertical (degr.)	21.54 (8.11)	43.08 (15.55)	47	−17.12	0.000 *	1.24

* ... significant at the Bonferroni-corrected (nominal) significance level.

Table 2. Descriptive statistics of visual behavior measures during the *take-overs* (TOR, manual driving, handback) in the conditions heads-up display (HUD) and auditory speech display (ASD). We report the mean (M) and standard deviation (SD), *t*-test statistics, and effect sizes (Cohen's *d*).

TAKE-OVERS							
		HUD		ASD			
	Variable	M (SD)	M (SD)	df	<i>t</i>	<i>p</i>	Cohen's <i>d</i>
Glances at RoadCenter	Mean Glance Dur. (ms)	10,484.69 (4346.67)	9782.86 (5630.95)	47	0.78	0.440	0.11
	Glance Rate (per sec.)	0.12 (0.04)	0.13 (0.04)	47	−1.82	0.075	0.26
	Glance Loc. Prob. (rel.)	0.61 (0.14)	0.73 (0.15)	47	−4.17	0.000 *	0.60
Visual Distraction	Perc. Transition Time (rel.)	0.06 (0.06)	0.06 (0.08)	47	−0.003	0.998	0.00
	PEORT	0.32 (0.13)	0.30 (0.12)	47	1.33	0.189	0.183
	PRC	0.63 (0.14)	0.66 (0.13)	47	−1.28	0.207	0.248
	PEGOR	0.577 (0.266)	0.177 (0.193)	47	8.13	0.000 *	−1.72
Link Value Prob. from/to RoadCenter	to/from HUD	0.10 (0.10)	-	-	-	-	-
	to/from Other	0.05 (0.07)	0.13 (0.14)	47	−4.42	0.000 *	0.63
	to/from Road	0.09 (0.10)	0.15 (0.14)	47	−2.48	0.017	0.36
	to/from DMD	0.0005 (0.003)	0.00 (0.00)	47	1.00	0.322	0.14
	to/from InstrumentCluster	0.02 (0.05)	0.02 (0.05)	47	0.60	0.554	0.09
Gaze Dispersion	Horizontal (degr.)	13.32 (2.35)	13.53 (2.09)	47	−0.83	0.413	0.12
	Vertical (degr.)	28.16 (14.80)	29.46 (12.47)	47	−0.68	0.501	0.02

* ... significant at the Bonferroni-corrected (nominal) significance level.

3.1. Visual Behavior

Eye-tracking recordings of 5 participants did not fulfill the necessary data quality standards (pupil recognition rate > 90 %) and hence had to be excluded. All measures were recorded per driving phase and averaged for the condition. We defined the measures glance rate, glance location probability, and link value probabilities as 0.00, in case of no glances at the respective AoI; in this case, mean glance duration was excluded from the hypothesis testing sample. Descriptive and test statistics are listed in Table 1 for the automated driving phases (cf. H1 and H2), and in Table 2 for the take-over phases including the TOR, manual driving, and the handback maneuver, respectively. A typical visual behavior is illustrated in Figure 4, showing all fixations of a participant in both conditions and driving phase types. All tests were performed on the average behavior during each driving phase type per condition and participant.

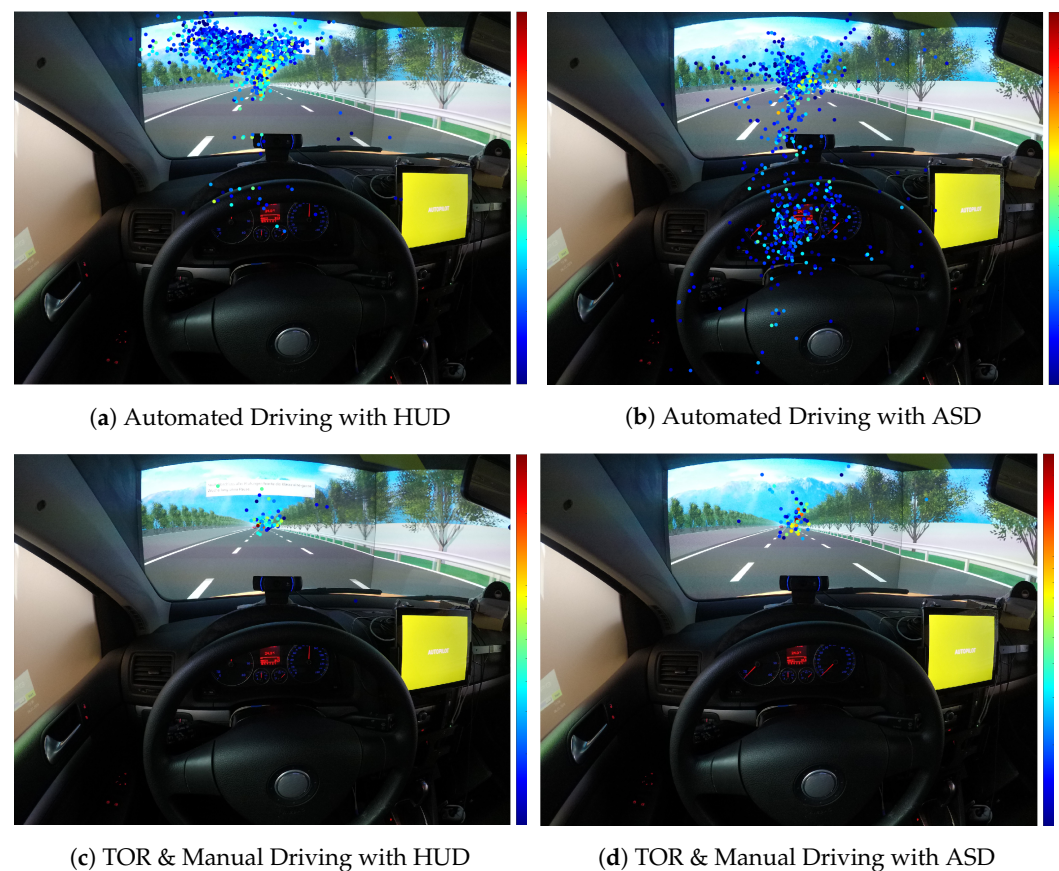


Figure 4. A participant's fixations during automated driving (a,b) and during take-over phases with manual driving and handback to to automated driving (c,d) per condition (HUD: a,c; ASD: b,d). The color scale indicates fixation duration with logarithmic scaling from the minimum (blue) to the maximum fixation duration (red).

3.1.1. Area of Interest-Specific Behavior

In automated driving, only four participants (out of 47) glanced at the DMD in both conditions (HUD: 7, ASD: 17). Thus, we excluded the AoI DMD from further analysis. A portion of participants ($n = 15$) never glanced at AoI InstrumentCluster in condition HUD ($n = 2$ in condition ASD) resulting in 33 complete samples for mean glance duration comparison. The mean glance duration and glance location probability at AoI RoadCenter were significantly higher in condition ASD. In contrast, glance rate at the RoadCenter was significantly lower than with the HUD. Concerning AoIs InstrumentCluster and "Other", no significant effect between conditions HUD and ASD was found for mean glance duration and glance rate, for either of both AoIs. However, we found statistically significant

differences with medium to large effect sizes for glance location probability at both AoIs. The probability to glance at InstrumentCluster and “Other” was ca. 11 % and 12 % higher on average, with the ASD than with the HUD.

In take-overs, we analyzed glance behavior only at AoI RoadCenter. All 47 participants glanced this AoI in both conditions. Hence, no samples are excluded from analysis in this phase. Statistically significant differences were found for the glance location probability, indicating to be, on average, ca. 12% higher with the ASD than with the HUD. Measures mean glance duration and glance rate did not show significant effects.

3.1.2. Visual Distraction

In automated driving, the percentage road center (PRC) was significantly higher with the ASD than with the HUD, with an average difference of ca. 28 %. Percentage eyes off road time (PEORT) comparisons naturally showed the inverse effect with a sufficiently large effect size. The time spent transitioning between AoIs was not significantly different in the comparison of conditions.

In take-overs, tests revealed no significant difference regarding PRC, PEORT and transition time. However, glances off the road with more than 2 s duration were over three times more likely in condition HUD, also highlighted by the sufficiently large effect size.

3.1.3. Transitioning Behavior and Gaze Dispersion

An illustration of average glance transitioning behavior can be found in Figure 5. A typical example of gaze dispersion is highlighted in Figure 4 using a fixation plot.

In automated driving, the probabilities for transitioning from/to any AoI (Other, Road, InstrumentCluster, DMD) to/from AoI RoadCenter—the Link Value Probabilities—were significantly higher in condition ASD, in comparison to condition HUD. Specifically, large effects could be observed in the link value probabilities to/from AoI “Other”. On average, a participant’s probability to glance to/from AoI “Other” was (on average) 42 % higher in condition ASD, than in condition HUD. The total time to perform these transitions, the percentage transition time, was not significantly different between conditions. Still, gaze dispersion analysis revealed statistically significant differences in both axes with considerably large effect sizes.

In take-overs, comparing link value probabilities from/to RoadCenter between the conditions reveals a significant difference for the transitioning probability RoadCenter $\Leftrightarrow$ “Other”. The medium effect showed a descriptive mean difference of ca. 8 %. RoadCenter $\Leftrightarrow$ DMD and RoadCenter $\Leftrightarrow$ InstrumentCluster showed no significant difference; RoadCenter $\Leftrightarrow$ Road missed the Bonferroni-corrected threshold ($\alpha' = 0.0125$) slightly. The percentage transition time, again, was not significantly different in the comparison of conditions. In contrast to automated driving, testing showed no significant differences in gaze dispersion.

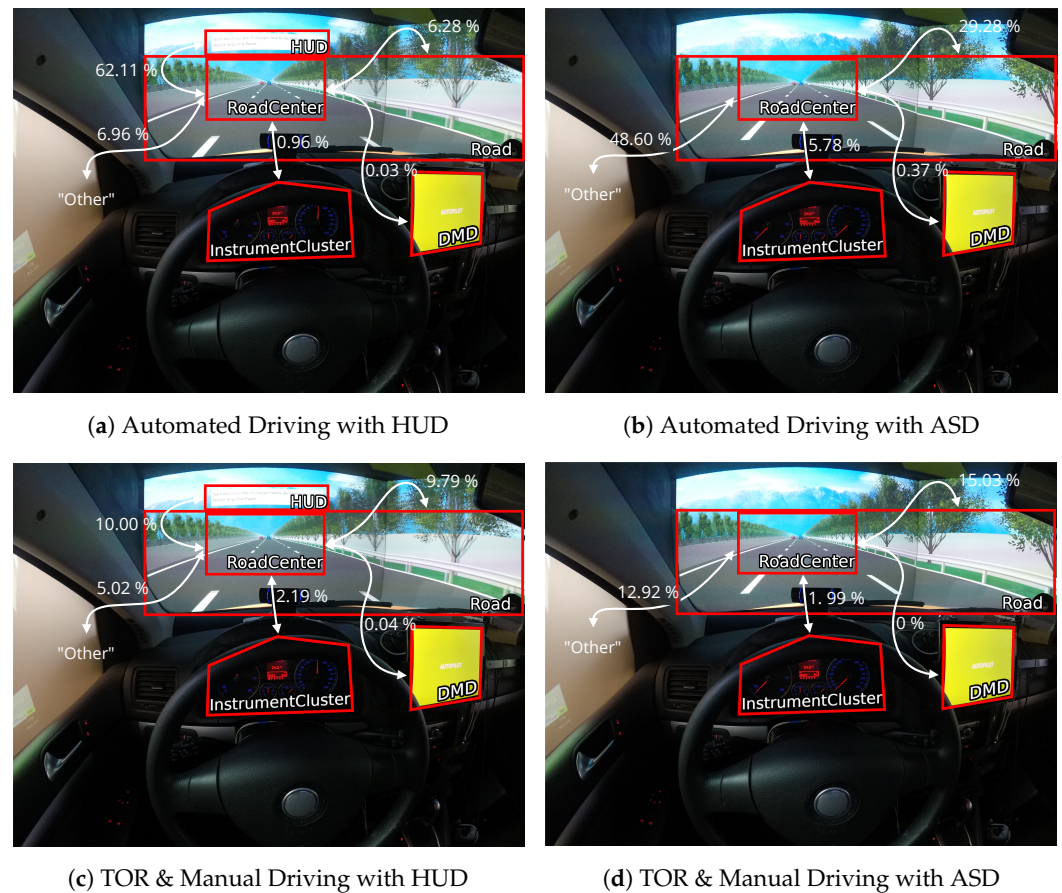


Figure 5. Mean link value probabilities between analyzed AoIs depicted as white double-headed arrows with probability labels: during automated driving (a,b) and during take-over phases with manual driving and handback to to automated driving (c,d) per condition (HUD, a,c; ASD, b,d). The HUD was not visible during take-overs.

3.2. Performance, Workload, and Acceptance

We were able to confirm our findings in [38] with the extended sample size ($N = 53$ instead of $N = 32$), using the same test methods: Driving performance (i.e., maneuvering quality after a TOR) was not significantly different between NDRT display conditions, participants were significantly faster to execute the NDRT in the HUD condition, and physiologically measured tonic (long-term) workload measurements were significantly lower in the HUD condition during automated driving. In contrast, enjoyability, indicated by technology acceptance and perceived workload self-ratings, favored the ASD.

Some minor differences in comparison to the first sample were observed though, in the extended sample: Concerning self-rated workload, items Mental Demand and Physical Demand did not reach the Bonferroni-corrected significance-threshold ($\alpha' = 0.0125$) anymore ($p_{Mental} = 0.017$, $p_{Physical} = 0.088$). Instead, Temporal Demand's ratings were significantly worse (higher) with the HUD than with the ASD: $Mdn_{HUD} = 4$, $Mdn_{ASD} = 3$; $Z = -2.77$, $p = 0.006$; Effect size $r = 0.27$. Take-over response comparisons of the time (in seconds) to first driving action (RTd) and time to first gaze on the road (RTe) now also revealed a significant main effect between conditions HUD and ASD, after Bonferroni correction ($\alpha' = 0.025$): $M(RTd)_{HUD} = 2.50$, $SE(RTd)_{HUD} = 0.077$; $M(RTd)_{ASD} = 2.639$, $SE(RTd)_{ASD} = 0.073$; $M(RTe)_{HUD} = 0.238$, $SE(RTe)_{HUD} = 0.029$; $M(RTe)_{ASD} = 0.109$, $SE(RTe)_{ASD} = 0.026$; $p(RTe) = 0.022$, $p(RTd) = 0.000$; Effect sizes $\eta_p^2(RTd) = 0.10$, $\eta_p^2(RTe) = 0.23$). They show that RTd was faster and RTe slower in condition HUD, in comparison to condition ASD.

4. Discussion

To answer our research questions, we studied visual attention during automated driving phases, where users engaged in a text comprehension NDRT, and during take-over phases, where users had to engage in manual driving upon a TOR until vehicle control could be given back to the automated driving system. Moreover, we updated our analysis of NDRT and take-over performance, self-ratings, and physiologically measured workload with the full sample, to jointly discuss the results with visual behavior as a possible cause of the previously investigated effects.

4.1. Visual Behavior in Automated Driving (RQ1, H1 and H2)

We found that, in accordance with hypothesis H1, the non-overlapping HUD has shown more frequent and shorter glances to the road center during automated driving in comparison to the ASD. This also supports Zeeb et al.'s findings [5] for NDRTs displayed in the center console. In our study, the mean difference of mean glance durations on the road center is about 10 s, which is similar to the mean glance duration on the HUD. One could thus assume that the “visual freedom” with the ASD would result in a nearly exclusive focus of visual attention on the road center. However, larger raw gaze dispersions in the ASD condition emphasize other areas also drew more attention, in comparison to the HUD condition. We further found statistically significant effects for glance location probability on AoI InstrumentCluster and link value probabilities between AoI RoadCenter and the other analyzed AoIs. These indicate that, relative to the number of AoIs available, glancing at other AoIs than the road center is more likely in condition ASD—AoI HUD, was, however, neither available nor relevant in this condition. One might argue that this is obvious due to the fact that there are more AoIs in the HUD condition (i.e., the HUD is added), and glance rates and mean glance durations also do not support this argument. However, we believe this matter is not that trivial, since, for example, executing the NDRT was significantly quicker with the HUD, and a lack of significance may also be caused by other factors (sample size, data distribution, ...)—in fact, these measures have not shown any statistically significant differences. We argue that participants could also have used the NDRT performance and therefore time-advantage of the HUD to glance at other AoIs than RoadCenter and HUD. The result would have been similar glancing behavior as with the ASD—but, as our results on the gaze dispersion measurements have also shown, visual attention was concentrated to a much more narrow area with the HUD. Interestingly, glances at AoI DMD (driving mode display) were so few in both conditions that we had to exclude the AoI from hypothesis testing in regards to glance analysis. Overall, we argue to accept hypothesis H1.

Concerning hypothesis H2, glance location probability on AoI “Other” is statistically significantly higher in condition ASD, with a sufficiently large effect and an average difference of ca. 12%. Further supporting H2 is the link value probability between AoIs RoadCenter and “Other”, indicating that users were about 42% more likely to/from AoI “Other”. Overall, we argue to partly accept hypothesis H2, stating that it is indeed more likely to glance at AoI “Other” relative to the available number of AoIs in condition ASD. However, overall frequency and mean glance duration at AoI “Other” have not shown the same effect.

In summary, we found strong evidence supporting the acceptance H1, stating that users spend more visual attention on the road center. We could further also show that with ASD users do not exclusively limit their visual attention to the center of the road, where an upcoming TOR could be anticipated, either. Raw gaze dispersion, glance location and transitioning probabilities, on the one hand, show that driving-related areas, with the exception of the driving mode display, are more likely to be glanced at with the ASD. On the other hand, visual transitions to areas that are not related to either of the primary tasks, are also more likely with the ASD than with the HUD, at least partly supporting H2.

4.2. Visual Behavior during Take-Overs (RQ1, H3)

Visual distraction measure analysis (PRC and PEORT) shows that during automated driving with the ASD, more time was spent looking at the road center but, at the same time, raw gaze was much more dispersed than with the HUD, potentially leading to increased awareness of surroundings. As a consequence, with the ASD, users should need to glance less extensively at other AoIs than with the HUD, in order to establish sufficient situation awareness after a TOR is issued, and therefore be able to more safely engage in manual driving.

With the HUD, visual attention is possibly too focused on the center windshield area during NDRT engagement, thereby leading to low maintained levels of situation awareness. Upon a TOR, this would naturally lead to an increased need to glance at other AoIs than at the road center directly (such as the instrument cluster or DMD). Our results partly support this assumption with a higher glance location probability on the road center in condition ASD but no significant results in regards to total times of gaze on/off road (PRC and PEORT). However, the PEGOR indicator has shown a significantly increased—over 3 times more likely on average—percentage of safety-critical off road glances with the HUD still. In the same manner, visual reaction times after a TOR (RTe) were significantly better in condition ASD (ca. 120 ms lower on average). Interestingly, raw gaze dispersion and link value probabilities were not significantly higher during take-overs in condition HUD, than in condition ASD. In contrast, the probability to transition between AoIs RoadCenter and “Other” was higher with the ASD by about 8% on average.

To summarize, in both conditions we found evidence that AoIs unrelated to the driving task were still glanced at after users received the TOR. However, in condition HUD, about 58% of glances not at the road could have been considered a potential safety risk (PEGOR), in contrast to condition ASD’s PEGOR of ca. 18%. This indicates an increased need to establish sufficient situation awareness in condition HUD after a TOR was received. Thus, we argue to accept H3.

4.3. Impact on Safety, Productivity and User Experience (RQ2)

Previously, we [38] found that ASDs cause significantly higher workload (physiologically measured) and worse productivity (NDRT performance) than HUDs during automated driving in our setting. At the same time, participants subjectively favored ASDs. Our extended sample could mostly confirm these findings (cf. Section 3.2) and additionally showed that there was a significant impact of NDRT display modality on the take-over reaction: the time to first manual driving action upon a TOR (RTd), as well as the time to first gaze on road (RTe)—both are considered an important predictor of take-over safety [5] but arguably not the only one, as maneuvering (i.e., driving) performance after the control transition has to be considered too [41], where we found no significant differences.

Interestingly, in the HUD condition, users were slower to gaze at the road than in the condition ASD, and had a significantly higher average percentage of extended glances on AoIs that do not involve the road. These results, where visual behavior was moderated by the two NDRT display variants, stand in contrast to Zeeb et al. [5], who argued that monitoring strategies do not affect the time to first glance on the road. We showed the differences in monitoring strategies during the automated driving phases in the evaluation of RQ1, finding that with a HUD for the NDRT, participants focused strongly on the center section of the windshield by mainly transitioning frequently between the HUD and road center and glancing significantly less likely at other AoIs, in comparison to when using the ASD. This restriction of visual attention may have led to extended glancing off road during the take-over maneuver since participants possibly needed to regain sufficient situational awareness of their environment and the vehicle’s state. The significantly “better” time to first driving action (RTd) in the HUD condition could thereby also simply be an indicator of stress reactions since the increased visual distraction and arguably thereby increased cognitive demand during the take-over phase with preceding HUD usage, in comparison to preceding ASD usage, is also reflected in the mean (absolute) pupil diameter change

(MPDC [57], in millimeters) from automated driving to the take-over maneuver: $M_{HUD} = 0.47$, $SD_{HUD} = 0.16$, $M_{ASD} = 0.25$, $SD_{ASD} = 0.16$; $t(47) = 7.31$, $p = 0.000$, $Cohen'sd = -1.37$). While the take-over phases were relatively short in comparison to the automated driving phases, they apparently were very arousing as shown by the MPDC measure, leading to a higher overall workload impression in the HUD condition, as indicated by participants' self-ratings. Overall, these high(er) demand measures during take-overs with the HUD have not caused a significant difference in the investigated driving performance measures (i.e., maneuvering quality after a TOR) and in total, every participant managed to complete every take-over maneuver. The Yerkes-Dodson law [58] would hence suggest that either the task (i.e., the TOR maneuver) was too simple to reveal differences or the workload was not high enough to impair performance. Considering our strict TOR lead time of 5 s and the unpredictable scenario of two out of three vehicles in front suddenly stopping, we do not believe that the TOR task was too simple. We rather believe the reason for that is that the text comprehension task was well manageable with the used interaction modalities and display variants, and thereby not overloading the user after all. Similarly, Borojeni et al. [59] found that text comprehension tasks of varying difficulty can be well-manageable in comparable situations, and the previously mentioned Engström et al.'s Cognitive Control Hypothesis [33] would also suggest potential disadvantages to diminish if the task and user are well-enough adapted.

4.4. Implications for Future Interfaces

Drivers of SAE L3 automated vehicles need to stay ready to take over while engaging in an NDRT. Perceiving the environment to maintain good situational awareness is a visually demanding task. Thus, applying the Multiple Resource Theory [32], we hypothesized that an auditory display for the NDRT would distract less than a visual display and help to stay ready to take over control. However, a visual display should be more efficient at presenting information and thus lead to improved NDRT performance. Yet it was unclear, whether optimizing a visual (otherwise traditionally “heads-down”) display to a HUD, could compensate the expected downsides of the visual NDRT display modality, in comparison to an auditory one. Our results have shown that the HUD still had a stronger effect on visual distraction and workload during take-overs than the ASD but ultimately not on driving performance. Hence, we cannot clearly recommend one NDRT display modality over the other. Looking into the trade-offs, we argue for using a speech interface though. While the drivers performed the NDRT slower (though not with significantly more errors) than in the HUD condition, they showed less risky visual behavior in take-overs. The ability to adjust speech output live and other more advanced functionalities may further help to speed up the NDRT task and thereby bring its productivity to the HUD's level. After all, participants preferred this auditory display, too.

While the ASD may be considered somewhat safer during take-overs, the HUD is still worth considering for NDRTs, especially since no safety-relevant driving performance measure besides visual behavior was affected significantly in our experiment. We would thus argue that a 5 s TOR lead time is also long enough when using a HUD, at least in our investigated road scenario. However, we expect that the difference between HUD and ASD will become more prominent the shorter the take-over time and the more complex the road scenario becomes. An interesting feature of the HUD is that it received most of the driver's attention, besides the very road center, during NDRT engagement. Riegler et al. [60] have used this aspect by dynamically positioning text on the windshield, near relevant objects in the driving scene. However, this would require an automated driving system to correctly identify such objects—then the question has to be asked, why is a TOR necessary in the first place? Future work could also investigate if it is beneficial to, for example, combine both driving-related information and the NDRT on the HUD. Still, drivers could be (cognitively) focused on the NDRT to an extent that makes it hard to perceive and process much else, even if more driving-related information is close.

While it was not necessarily part of our research question, we also noticed that our participants did not pay much direct attention to the driving mode display. Following the credo that an invisible interface is also not usable, we, for example, suggest placing the screen closer to the instrument cluster. Considering that a driver's visual focus is mainly on the street during manual driving, it may be even more beneficial to actively address peripheral vision using an ambient light display (cf., [61,62]) or another display modality that does not address a cognitive resource that is already loaded by the maintaining situational awareness and possibly performing the NDRT. The design of this interface may impact the performance when handing back driving control to the automation and thus help optimizing safety.

4.5. Limitations

Our visual behavior evaluation is based on a head-mounted eye-tracker with limited mapping capacities. This setup did not allow us to include AoIs at extreme angles (e.g., mirrors, side windows). To account for that, we could only remove the possibility for upcoming traffic, possibly overtaking the ego-vehicle, being relevant in the take-over situation. Additionally, in contrast to real automated driving with acceptable reliability, TORs occurred very frequently (every few minutes) which may have prevented habituation effects such as boredom or sleepiness, or even exhaustion due to the NDRT. In a similar manner, the driving situation and NDRT in the driving simulator was artificially fabricated to be comparable, but could thereby also have reduced immersion.

4.6. Future Work

In future research, we plan to explore the options of combining driving-related and unrelated information on a HUD display and its impact on situation awareness and workload, as well as more advanced technological implementations of HUDs (e.g., near-eye [63] or adaptive in display-depth [60,64]). Additionally, we want to investigate options to improve the auditory display's productivity potential by introducing options to live-adjust pacing of output. In regards to methodology, the results need to be confirmed in higher fidelity testing, for example, test-track or field operational tests, as soon as safe testing is possible. Further parameters have to be investigated, such as user characteristics (e.g., different age ranges, inclusion of various disabilities, ...), take-over scenarios, extended TOR lead times, different NDRT types, etc. Moreover, we intend to investigate the effects of longer-term engagements—for example, mind wandering and (mental) exhaustion.

5. Conclusions

In this article, we investigated two advanced display variants for NDRT engagement in conditionally automated vehicles (SAE L3). Their existing implementations and related research as driving-related information displays would suggest so, but an evaluation of their usage as NDRT displays was missing so far. Especially in a “mobile office” scenario, where complex and demanding NDRTs are not only required to be productive but also safe, research is needed that investigates realistic designs and tasks [65]. We, therefore, compared a Heads-Up Display (HUD) that was displayed semitransparent above the road center on the windshield to a text-to-speech engine, that acted as an Auditory Speech Display (ASD) in a comprehensive within-subjects driving simulator user study. In accordance to the Multiple Resource Theory [32], the auditory speech display was assumed to be advantageous, since the visual demand of the NDRT is shifted to the auditory perception channel, leaving the visual channel free for the task of anticipating upcoming TOR scenarios. The diverse results of our study suggest a more complex picture:

- Users subjectively favored the ASD over the HUD, despite higher overall physiologically measured workload and worse productivity.
- HUD usage during automated driving has a distinctly negative effect on visual behavior in the take-over phase too and results in a heavier situational workload increase than the ASD.

- Driving performance measures of manual driving were not impacted in a statistically significant fashion. Hence, we assume that both NDRT display variants are still safe to use, even though the users reacted quicker to TORs in the ASD condition.

Overall, we suggest to:

- use Auditory Speech Displays for productivity in conditionally automated vehicles (SAE L3) to maximize safety and enjoyability (acceptance and perceived workload). However, research and design should consider interaction improvement options to alleviate its productivity downsides in terms of efficiency, such as live adjustability of output parameters (e.g., pacing, rhythm).
- consider combining driving-related information with NDRT-information on HUDs to potentially alleviate its downsides in terms of situation awareness.

Author Contributions: Conceptualization, C.S., K.W., P.W.; methodology, all authors; software, C.S. and P.W.; validation, all authors; formal analysis, C.S. and K.W.; investigation, C.S. and K.W.; resources, K.W., M.S., and A.R.; data curation, C.S., K.W., and P.W.; writing—original draft preparation, C.S., K.W., A.L., P.W.; writing—review and editing, all authors; visualization, C.S., K.W., A.L., P.W.; supervision, M.S. and A.R.; project administration, K.W. and M.S.; funding acquisition, K.W., M.S., A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This project was supported by the “Innovative Hochschule” program of the German Federal Ministry of Education and Research (BMBF) under Grant No. 03IHS109A (MenschINBewegung) and the FH-Impuls program under Grant No. 13FH7I01IA (SAFIR), as well as the German Federal Ministry of Transport and Digital Infrastructure (BMVI) through the *Automated and Connected Driving* funding program under Grant No. 16AVF2145F (SAVe). Participant compensation was funded by the proFOR+ program for research projects of the Catholic University of Eichstätt-Ingolstadt.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the ethics committee of the Catholic University of Eichstätt-Ingolstadt (protocol number: 2019/05; date of approval: 16 January 2019).

Informed Consent Statement: Informed consent was obtained from all subjects prior to the beginning of the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We want to thank our deceased lab engineer Stefan Cor for his support in conducting the study. His memory will be with us.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. SAE. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*; J3016 Ground Vehicle Standard; SAE International: Warrendale, PA, USA, 2018; doi:10.4271/J3016_201806.
2. McDonald, A.D.; Alambeigi, H.; Engström, J.; Gustav Markkula, G.; Vogelphl, T.; Dunne, J.; Yuma, N. Toward Computational Simulations of Behavior During Automated Driving Takeovers: A Review of the Empirical and Modeling Literatures. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **2019**, *61*, 642–688, doi:10.1177/0018720819829572.
3. Zhang, B.; de Winter, J.; Varotto, S.; Happee, R.; Martens, M. Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *64*, 285–307.
4. Gold, C.; Berisha, I.; Bengler, K. Utilization of Drivetime—Performing Non-Driving Related Tasks While Driving Highly Automated. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2015**, *59*, 1666–1670, doi:10.1177/1541931215591360.
5. Zeeb, K.; Buchner, A.; Schrauf, M. What determines the take-over time? An integrated model approach of driver take-over after automated driving. *Accid. Anal. Prev.* **2015**, *78*, 212–221, doi:10.1016/j.aap.2015.02.023.
6. Young, M.S.; Stanton, N.A. Attention and automation: New perspectives on mental underload and performance. *Theor. Issues Ergon. Sci.* **2002**, *3*, 178–194, doi:10.1080/14639220210123789.
7. Feldhütter, A.; Kroll, D. Wake Up and Take Over! The Effect of Fatigue on the Take-over Performance in Conditionally Automated Driving. In Proceedings of the IEEE 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018.

8. Weinbeer, V.; Muhr, T.; Bengler, K.; Ag, A. Automated Driving: The Potential of Non-driving-Related Tasks to Manage Driver Drowsiness. In Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018), Florence, Italy, 26–30 August 2018; Springer International Publishing: Cham, Switzerland, 2019; Volume 827, pp. 179–188, doi:10.1007/978-3-319-96059-3.
9. Miller, D.; Sun, A.; Johns, M.; Ive, H.; Sirkin, D.; Aich, S.; Ju, W. Distraction Becomes Engagement in Automated Driving. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2015**, *59*, 1676–1680, doi:10.1177/1541931215591362.
10. Hecht, T.; Feldhütter, A.; Draeger, K.; Bengler, K. What Do You Do? An Analysis of Non-driving Related Activities During a 60 Minutes Conditionally Automated Highway Drive. In Proceedings of the International Conference on Human Interaction and Emerging Technologies, Nice, France, 22–24 August 2019; Springer: Cham, Switzerland, 2020; pp. 28–34, doi:10.1007/978-3-030-25629-6_5.
11. Pfleging, B.; Rang, M.; Broy, N. Investigating User Needs for Non-Driving-Related Activities During Automated Driving. In Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia, Rovaniemi, Finland, 12–15 December 2016; pp. 91–99, doi:10.1145/3012709.3012735.
12. WHO. Global Status Report on Road Safety 2015; Technical Report; WHO: Geneva, Switzerland, 2015.
13. Häuslschmid, R.; Pfleging, B.; Butz, A. The Influence of Non-driving-Related Activities on the Driver's Resources and Performance. In *Automotive User Interfaces: Creating Interactive Experiences in the Car*; Meixner, G., Müller, C., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 215–247, doi:10.1007/978-3-319-49448-7_8.
14. Wintersberger, P. Automated Driving: Towards Trustworthy and Safe Human-Machine Cooperation. Ph.D. Thesis, Universität Linz, Linz, Austria, 2020.
15. Perterer, N.; Moser, C.; Meschtscherjakov, A.; Krischkowsky, A.; Tscheligi, M. Activities and Technology Usage while Driving. In Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16), Gothenburg, Sweden, 23–27 October 2016; ACM Press: New York, NY, USA, 2016; pp. 1–10, doi:10.1145/2971485.2971556.
16. Hensch, A.C.; Rauh, N.; Schmidt, C.; Hergeth, S.; Naujoks, F.; Krems, J.F.; Keinath, A. Effects of secondary Tasks and Display Position on Glance Behavior during partially automated Driving. In Proceedings of the 6th Humanist Conference, The Hague, The Netherlands, 13–14 June 2018.
17. Wintersberger, P.; Schartmüller, C.; Riener, A. Attentive User Interfaces to Improve Multitasking and Take-Over Performance in Automated Driving. *Int. J. Mob. Hum. Comput. Interact.* **2019**, *11*, 40–58, doi:10.4018/IJMHCI.2019070103.
18. Li, X.; Schroeter, R.; Rakotonirainy, A.; Kuo, J.; Lenné, M.G. Effects of different non-driving-related-task display modes on drivers' eye-movement patterns during take-over in an automated vehicle. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *70*, 135–148, doi:10.1016/j.trf.2020.03.001.
19. von Sawitzky, T.; Wintersberger, P.; Riener, A.; Gabbard, J.L. Increasing Trust in Fully Automated Driving: Route Indication on an Augmented Reality Head-up Display. In Proceedings of the 8th ACM International Symposium on Pervasive Displays (PerDis '19), Palermo, Italy, 12–14 June 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–7, doi:10.1145/3321335.3324947.
20. Wintersberger, P.; Frison, A.K.; Riener, A.; Sawitzky, T.v. Fostering user acceptance and trust in fully automated vehicles: Evaluating the potential of augmented reality. *PRESENCE: Virtual Augment. Real.* **2019**, *27*, 46–62.
21. Riegler, A.; Wintersberger, P.; Riener, A.; Holzmann, C. Augmented Reality Windshield Displays and Their Potential to Enhance User Experience in Automated Driving. *i-com* **2019**, *18*, 127–149, doi:10.1515/icom-2018-0033.
22. Edgar, G.K.; Pope, J.C.; Craig, I.R. Visual accommodation problems with head-up and helmet-mounted displays? *Displays* **1994**, *15*, 68–75, doi:10.1016/0141-9382(94)90059-0.
23. Gang, N.; Sibi, S.; Michon, R.; Mok, B.; Chafe, C.; Ju, W. Don't Be Alarmed: Sonifying Autonomous Vehicle Perception to Increase Situation Awareness. In Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Toronto, ON, Canada, 23–25 September 2018; pp. 237–246.
24. Walker, B.N.; Nance, A.; Lindsay, J. *Spearcons: Speech-Based Earcons Improve Navigation Performance in Auditory Menus*; Georgia Institute of Technology: Atlanta, GA, USA, 2006.
25. Jeon, M. Lyricons (Lyrics+ Earcons): Designing a new auditory cue combining speech and sounds. In Proceedings of the International Conference on Human-Computer Interaction, Las Vegas, NV, USA, 21–26 July 2013; Springer: Berlin/Heidelberg, Germany, 2013; pp. 342–346.
26. Stratmann, T.C.; Löcken, A.; Gruenefeld, U.; Heuten, W.; Boll, S. Exploring Vibrotactile and Peripheral Cues for Spatial Attention Guidance. In Proceedings of the 7th ACM International Symposium on Pervasive Displays (PerDis '18), Munich, Germany, 6–8 June 2018; ACM: New York, NY, USA, 2018; pp. 9:1–9:8, doi:10.1145/3205873.3205874.
27. Petermeijer, S.M.; Abbink, D.A.; Mulder, M.; de Winter, J.C.F. The Effect of Haptic Support Systems on Driver Performance: A Literature Survey. *IEEE Trans. Haptics* **2015**, *8*, 467–479, doi:10.1109/TOH.2015.2437871.
28. Rose, D. Braille Is Spreading But Who's Using it? *BBC News*, 2012. Available online: <https://www.bbc.com/news/magazine-16984742> (accessed on 8 February 2021).
29. Large, D.R.; Banks, V.A.; Burnett, G.; Baverstock, S.; Skrypchuk, L. Exploring the behaviour of distracted drivers during different levels of automation in driving. In Proceedings of the 5th International Conference on Driver Distraction and Inattention (DDI2017), Paris, France, 20–22 March 2017; pp. 20–22.

30. You, F.; Wang, Y.; Wang, J.; Zhu, X.; Hansen, P. Take-Over Requests Analysis in Conditional Automated Driving and Driver Visual Research Under Encountering Road Hazard of Highway. In *Advances in Human Factors and Systems Interaction*; Nunes, I.L., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 230–240.
31. Chang, C.C.; Sodnik, J.; Boyle, L.N. Don't Speak and Drive: Cognitive Workload of In-Vehicle Speech Interactions. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '16 Adjunct), Ann Arbor, MI, USA, 24–26 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 99–104, doi:10.1145/3004323.3004351.
32. Wickens, C.D. Processing Resources in Attention, Dual Task Performance and Workload Assessment; In *Varieties of Attention*; Parasuraman, R., Davies, D.R., Eds.; Academic Press: Orlando, FL, USA, 1984; pp. 63–102.
33. Engström, J.; Markkula, G.; Victor, T.; Merat, N. Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Hum. Factors: J. Hum. Factors Ergon. Soc.* **2017**, *59*, 734–764, doi:10.1177/0018720817690639.
34. Endsley, M.R. Situation awareness misconceptions and misunderstandings. *J. Cogn. Eng. Decis. Mak.* **2015**, *9*, 4–32.
35. Moore, K.; Gugerty, L. Development of a novel measure of situation awareness: The case for eye movement analysis. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2010**, *54*, 1650–1654.
36. van de Merwe, K.; van Dijk, H.; Zon, R. Eye movements as an indicator of situation awareness in a flight simulator experiment. *Int. J. Aviat. Psychol.* **2012**, *22*, 78–95.
37. Zhang, T.; Yang, J.; Liang, N.; Pitts, B.J.; Prakah-Asante, K.O.; Curry, R.; Duerstock, B.S.; Wachs, J.P.; Yu, D. Physiological Measurements of Situation Awareness: A Systematic Review. *Hum. Factors* **2020**, doi:10.1177/0018720820969071.
38. Schartmüller, C.; Weigl, K.; Wintersberger, P.; Riener, A.; Steinhäuser, M. Text Comprehension: Heads-Up vs. Auditory Displays: Implications for a Productive Work Environment in SAE Level 3 Automated Vehicles. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '19), Utrecht, The Netherlands, 22–25 September 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 342–354, doi:10.1145/3342197.3344547.
39. Schartmüller, C.; Wintersberger, P.; Riener, A. Interactive Demo: Rapid, Live Data Supported Prototyping with U.S.E. In Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Oldenburg, Germany, 24–27 September 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 239–241, doi:10.1145/3131726.3132046.
40. Tsimhoni, O.; Green, P.; Watanabe, H. Detecting and Reading Text on HUDs: Effects of Driving Workload and Message Location. In Proceedings of the 11th Annual ITS America Meeting, Miami Beach, FL, USA, 4–7 June 2001.
41. Wintersberger, P.; Green, P.; Riener, A. Am I Driving or Are You or We Both? A Taxonomy For Handover and Handback in Automated Driving. In Proceedings of the 9th International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design, Manchester Village, VT, USA, 26–29 June 2017; Public Policy Center, University of Iowa: Iowa City, IA, USA, 2017; pp. 333–339.
42. Daneman, M.; Carpenter, P.A. Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* **1980**, *19*, 450–466, doi:10.1016/S0022-5371(80)90312-6.
43. Tobii, A. Tobii Pro Eye-Tracking-Brille 2. 2019. Available online: <https://www.tobii.com/de/produkte/tobii-pro-glasses-2/> (accessed on 9 September 2019).
44. iMotions, A/S. iMotions. 2019. Available online: <https://imotions.com/> (accessed on 13 March 2019).
45. Mole, C.D.; Lappi, O.; Giles, O.; Markkula, G.; Mars, F.; Wilkie, R.M. Getting Back Into the Loop: The Perceptual-Motor Determinants of Successful Transitions out of Automated Driving. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2019**, 1037–1065, doi:10.1177/0018720819829594.
46. ISO. *Road vehicles—Measurement of Driver Visual Behaviour with Respect to Transport Information and Control Systems—Part 1: Definitions and Parameters (ISO15007-1:2014)*; Technical Report; International Organization for Standardization: Geneva, Switzerland, 2015.
47. Louw, T.; Merat, N. Are you in the loop? Using gaze dispersion to understand driver visual attention during vehicle automation. *Transp. Res. Part C: Emerg. Technol.* **2017**, *76*, 35–50, doi:10.1016/j.trc.2017.01.001.
48. Victor, T.W.; Harbluk, J.L.; Engström, J.A. Sensitivity of eye-movement measures to in-vehicle task difficulty. *Transp. Res. Part F Traffic Psychol. Behav.* **2005**, *8*, 167–190, doi:10.1016/j.trf.2005.04.014.
49. Green, P. *Visual and Task Demands of Driver Information Systems*; Technical Report UMTRI-98-16; The University of Michigan Transportation Research Institute: Ann Arbor, MI, USA, 1999.
50. Driver Metrics, Performance, Behaviors and States Committee. *Operational Definitions of Driving Performance Measures and Statistics*; Technical Report J2944; SAE International: Warrendale, PA, USA, 2015.
51. ISO. *ISO 9241-11 Ergonomics of Human-System Interaction—Part 11: Usability: Definitions and Concepts*; Technical Report; International Organization for Standardization: Geneva, Switzerland, 2018.
52. Hart, S.G.; Staveland, L.E. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Adv. Psychol.* **1988**, *52*, 139–183, doi:10.1016/S0166-4115(08)62386-9.
53. Lykken, D.T.; Venables, P.H. Direct measurement of skin conductance: A proposal for standardization. *Psychophysiology* **1971**, *8*, 656–672, doi:10.1111/j.1469-8986.1971.tb00501.x.

-
54. Davis, F.D. A Technology Acceptance Model for Empirically Testing New End-User Information Systems: Theory and Results. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 1985.
 55. IBM Corp. *IBM SPSS Statistics for Windows*; Version 25.0; IBM Corp.: Armonk, NY, USA, 2017.
 56. Cohen, J. A power primer. *Psychol. Bull.* **1992**, *112*, 155–159.
 57. Palinko, O.; Kun, A.L.; Shyrov, A.; Heeman, P. Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, Austin, TX, USA, 22–24 March 2010; pp. 141–144, doi:10.1145/1743666.1743701.
 58. Yerkes, R.M.; Dodson, J.D. The relation of strength of stimulus to rapidity of habit-formation. *J. Comp. Neurol. Psychol.* **1908**, *18*, 459–482.
 59. Borojeni, S.S.; Weber, L.; Heuten, W.; Boll, S. From reading to driving: Priming mobile users for take-over situations in highly automated driving. In Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '18), Barcelona, Spain, 3–6 September 2018; pp. 1–12, doi:10.1145/3229434.3229464.
 60. Riegler, A.; Riener, A.; Holzmann, C. Towards Dynamic Positioning of Text Content on a Windshield Display for Automated Driving. In Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology (VRST '19), Parramatta, Australia, 12–15 November 2019; Association for Computing Machinery: New York, NY, USA, 2019; doi:10.1145/3359996.3364757.
 61. Löcken, A.; Heuten, W.; Boll, S. Enlightening Drivers: A Survey on In-Vehicle Light Displays. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 97–104, doi:10.1145/3003715.3005416.
 62. Löcken, A.; Frison, A.K.; Fahn, V.; Kreppold, D.; Götz, M.; Riener, A. Increasing User Experience and Trust in Automated Vehicles via an Ambient Light Display. In Proceedings of the 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '20), Oldenburg, Germany, 5–9 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–10, doi:10.1145/3379503.3403567.
 63. Koulouris, G.A.; Akşit, K.; Stengel, M.; Mantiuk, R.K.; Mania, K.; Richardt, C. Near-eye display and tracking technologies for virtual and augmented reality. *Comput. Graph. Forum* **2019**, *38*, 493–519.
 64. Li, K.; Geng, Y.; Özgür Yöntem, A.; Chu, D.; Meijering, V.; Dias, E.; Skrypchuk, L. Head-up display with dynamic depth-variable viewing effect. *Optik* **2020**, *221*, 165319, doi:10.1016/j.ijleo.2020.165319.
 65. Chuang, L.L.; Donker, S.F.; Kun, A.L.; Janssen, C. Workshop on The Mobile Office. In Proceedings of the 10th International Conference on Automotive User Interfaces and Applications, Toronto, ON, Canada, 23–25 September 2018; doi:10.1145/3239092.3239094.