

ORIGINAL PAPER

Open Access



# Effects of non-driving related tasks on mental workload and take-over times during conditional automated driving

Andreas Lars Müller, Natacha Fernandes-Estrela, Ruben Hetfleisch, Lukas Zecha and Bettina Abendroth\* 

## Abstract

**Background:** Automated driving will be of high value in the future. While in partial-automated driving the driver must always monitor the traffic situation, a paradigm shift is taking place in the case of conditional automated driving (Level 3 according to SAE). From this level of automation onwards, the vehicle user is released from permanent vehicle control and environmental monitoring and is allowed to engage in Non-Driving Related Tasks (NDRT) in his or her newly gained spare time. These tasks can be performed until a take-over request informs the user to resume vehicle control. As the driver is still considered to be the fall-back level, this aspect of taking over control is considered especially critical.

**Methods:** While previous research projects have focused their studies on the factors influencing the take-over request, this paper focuses on the effects of NDRT on the user of the vehicle during conditional automated driving, especially on the human workload. NDRT (such as Reading, Listening, Watching a movie, Texting and Monitoring ride) were examined within a static driving simulator at the Institute of Ergonomics & Human Factors with 56 participants in an urban environment. These NDRT were tested for mental workload and the ability to take over in a critical situation. To determine the perceived workload, the subjective workload, psychophysiological activity as well as performance-based parameters of a secondary competing task performed by a were used.

**Results:** This study revealed that the selected NDRT vary significantly in their mental workload and that the workload correlates with the length of the time needed for take over control. NDRT which are associated with a high workload (such as Reading or Texting) also lead to longer reaction times.

**Keywords:** Automated driving, Mental workload, Take-over time, Driving simulator, Non-driving related task

## 1 Introduction

### 1.1 Motivation

The demand for individual passenger transport has increased considerably in recent decades, which has had not only positive but also negative side effects. The increase in traffic density resulted, among other things, in 2.6 million accidents on German roads in 2018, with around 400,000 people injured and approximately 3300 traffic fatalities. According to studies, around 86% of all

accidents involving personal injury are attributable to driver misconduct [1]. According to [2], even 95% of all fatal accidents are caused by human error. To counteract this, great expectations are therefore placed in automated driving. To reduce the number of accidents, automated driving systems can be used to protect the driver in complex driving situations from being overloaded or, in the case of reduced attention, from impending accidents. For many drivers, delegating the driving task is therefore also associated with a gain in comfort [3]. In addition, automated vehicles can potentially improve safety, reduce congestion and thus emissions, and positively influence the independence and mobility of the

\* Correspondence: [abendroth@iad.tu-darmstadt.de](mailto:abendroth@iad.tu-darmstadt.de)  
Technical University of Darmstadt, Department of Mechanical Engineering,  
Institute of Ergonomics and Human Factors, Otto-Berndt-Straße 2, 264287  
Darmstadt, Germany

non-driving population [4]. In the case of L3 driving [5], the driver can relinquish control of the vehicle. This makes it possible to deal with non-driving related tasks (NDRT) while driving. Tasks vary in their type and complexity and therefore require a different level of attention. First studies already show a reduced ability to take over as a result of performing NDRT during automated driving [6, 7]. The performed NDRT and the change of tasks may cause a reduction in the take-over control capability.

### 1.2 Scope of this paper

In this paper, the effects of different NDRT on the vehicle user during conditional automated drive are experimentally investigated. Since the vehicle user still serves as a fallback level during conditional automated driving, the aspect of the change of tasks from NDRT to taking over control is considered critical. The mental workload was investigated by means of psychophysiological and performance-based parameters as well as the subjective task load. Since in conditional automated systems it can happen that the vehicle user has to resume control of the vehicle guidance, this aspect will be investigated for different NDRT. In particular, the ability to take over after a take-over request (TOR) in urban traffic as well as the relationship between workload and take-over control will be analysed in this paper.

## 2 Theoretical principles

In the context of conditional automated driving, the vehicle user can turn away from the obligation of permanent vehicle control as well as monitoring the environment and engage in NDRT in his or her newly acquired spare time. These activities may be carried out until a TOR advises the user to resume vehicle control. Since the driver is the fallback level, this aspect of taking control is considered particularly critical as a late reaction of the driver to a TOR could result in accidents.

[8] have already proposed various interacting determinants and their implications for automated systems. For example, trust, mental models, experiences, task loads, situation awareness and mental workload should be used to explain behaviour during automation. According to [8–11] mental workload is a construct for explaining performance and safety in automated systems and is therefore described in more detail. In addition, aspects of take-over request, further literature references to previous research results and the research questions are presented below.

### 2.1 Mental workload

In order to understand the term workload, the stress and strain concept (SSC) (cf. [12, 13]) is briefly explained. A simple approach to explain this concept is the

cause-effect chain. The stresses are generally causes that are independent of the individual and have an effect on humans. Humans react to this in the form of quantifiable individual strain. In contrast, the workload concept is described as “portion of the operator’s limited capacity that is actually required to perform a particular task” [14]. According to [15] workload is understood as “[...] the specification of the amount of information processing capacity that is used for task performance” [16]. use the term workload to answer questions such as “How busy is the operator?“, “How complex are the tasks that the operator is required to perform?“, “Can any additional tasks be handled above and beyond those that are already imposed?“ [17]. defines workload as the ratio between the resources required by a task and the resources available to the human. According to [18], emotional and mental load is summarized as psychological load. Furthermore, emotional strain is often seen in direct connection with feelings. Mental workload, on the other hand, describes the cognitive reaction of the human information processing system to the informational parts [19]. It can be summarised from the above definitions that the stresses affecting humans result in strain respectively workload. The acting stress can be differentiated into task- and situation-specific partial stress. Partial stress that affects humans can be summarised to a total stress and cause measurable strain or workload in humans. The strain or workload is therefore the effect or the reaction of a person to external stress factors.

### 2.2 Take-over request

A central problem in conditionally automated vehicle research is how quickly the vehicle user can react to a critical event or a TOR. Until automated systems are able to perform all driving tasks under all conditions, the vehicle users must regain control if the automation fails or reaches its operating limits. Partial automation (L2), which is already provided by several car manufacturers, requires that the vehicle users constantly monitor the road and are able to intervene in case of critical events. In L3 vehicle users can delegate the monitoring task to the system during automated driving and therefore engage in NDRT.

The take-over process was previously described by [20]. The transfer process starts with the conditional automated vehicle guidance. If the automated system issues a TOR, it is necessary for the user to detect and register it. Then the change of tasks to vehicle takeover and guidance takes place by interrupting the NDRT that has been carried out and turning one’s gaze back to the road, before a choice of action is made. In parallel, the motor readiness is established. This is characterized by gripping the steering wheel with the hands and/or moving the feet to the pedals. Finally, manual control of the

vehicle can be taken over by steering and/or braking. How long the transition to manual driving takes and which factors explain the transfer time has already been investigated in recent years. The reaction time most commonly used in the scientific literature is the take-over time (TOT). It is defined as the time between TOR and the intervention in the vehicle control. This time already shows a wide bandwidth in the publications. In [21] an average brake reaction time of only 0.87 s is found, in a meta-analysis of 25 studies by [22] an average TOT of 2.97 s [min = 1.9 s; max = 25.7 s] is shown and at [23] a TOT of 3.2 s [max = 8.8 s] is observed.

### 2.3 Related work

The human-related research on conditional automated driving is primarily concerned with the question of how much time the driver needs to intervene in the driving task again. According to a meta-study by [24] 129 studies have been identified to determine the factors influencing the take-over time. Further analyses of previous studies on transition are provided by [25, 26]. In the literature reviews cited above, influencing factors such as urgency, environmental factors (including the complexity of the traffic situation) and the effect of NDRT are particularly mentioned.

Numerous studies have investigated the urgency of a takeover situation depending on the time available until a collision is impending, also called time budget or time-to-collision (TTC) [27], examined various time budgets and found that in more critical takeover situations (lower time budget) the reaction times were faster than in more extensive time horizons. The authors found that from a time budget of 6 to 8 s, there were no differences in the frequency of take-over control errors. In addition, [28] examined the effects of the time budget. Longer time budgets also lead to longer TOT.

The environmental factors, in particular the complexity of the traffic situation, were investigated at [29] as well as [30]. It turned out that a more complex traffic situation leads to longer TOT. However, this negative effect could not be found in [31].

For investigation in the driving context, the literature also contains a classification into standardised and more naturalistic NDRT. Standardised techniques intended to imitate more naturalistic NDRT are, for example, the cognitive loading n-Back Task [32] or the visual search task SuRT [33]. A list of standardised and naturalistic NDRT studies in the context of different degrees of automation can be found in [6]. Standardised tests have advantages such as better comparability and repeatability. The disadvantage of standardised tests can be seen as the lack of transferability of results to reality. Similarly, the motivation to perform tasks is supposedly higher in more naturalistic NDRT than in standardised

activities, which can lengthen the time needed to take over control. In [34] test persons performed the visually distracting SuRT and needed more time for a takeover than drivers without NDRT. Studies by [35] as well as [36] also used SuRT as a distracting activity and delivered similar results [29]. compared the effects of different NDRTs by means of SuRT and an n-back test on the ability to take-over vehicle control. The two NDRTs did not show significant differences in driving behaviour during the take-over situation. In the study by [37] a standardised quiz was used as a NDRT. The subjects did not react significantly different compared to a control group without additional activity. However, they showed a shorter time gap to an obstacle after taking the quiz.

Other studies focused on more naturalistic tasks such as reading news articles [38] [23]. investigated the different emphasis of NDRT in automated driving. In one experiment, several versions of a quiz game were implemented to simulate an increasing workload. In all versions, the question was played audibly, but the answer options were presented differently (acoustically or visually). The answering modalities were also varied (verbal or motor). The greatest impairment of acquisition ability was found for the variant that included a combination of acoustic, cognitive, visual and motor load. In a study by [25], participants performed two NDRTs on a tablet (reading a newspaper article, playing Tetris) and compared both NDRTs with a baseline test. In comparison to a control group, the takeover times for both NDRTs were significantly longer. However, the comparison among the NDRTs showed no significant difference.

The influence of different writing activities on a mobile device (texting) regarding the take-over quality during automated driving was investigated by [7]. They concluded that the different task modalities have an influence on the take-over quality. A motor-visual task (texting) shows worse reaction times than other NDRTs (visual-verbal) and when driving without NDRT [39]. also examined the influence of naturalistic NDRT (writing email, reading news and watching video) on takeover performance. No significant effects on reaction times (hand to the steering wheel) were found within the NDRTs investigated.

In this context, it can be concluded that different factors influencing the ability to take over during automated driving have already been identified and researched in the literature. Furthermore, it can be reported that standardised and naturalistic NDRT have already been investigated. However, comparatively few studies investigated more than just one NDRT. In addition, the research has shown that when several NDRTs with different demands were studied, no significant differences in ability to take over control were

found among the different activities depending on the study.

## 2.4 Research questions

The investigation of the workload caused by various naturalistic NDRTs during automated driving has not yet been sufficiently investigated and thus represents a research demand. In this paper the NDRT is considered as an independent object of investigation during automated driving, which results in the following research question:

### 2.4.1 RQ1: *how does the mental workload differ when performing different naturalistic NDRT during conditional automated driving?*

Since there is no explicit research on this issue, the following undirected difference hypothesis is made  $H_1$ : *There is a significant difference in the mental workload when performing different naturalistic NDRT during automated driving.*

So far, the reviewed studies indicate that NDRT have an impact on the take-over ability of vehicle users. How different naturalistic NDRT affect the ability to take over and whether this can be explained by the previously investigated construct mental workload is to be examined more closely with the second research question.

### 2.4.2 RQ2: *how does the take-over time differ between different naturalistic NDRTs and can this be explained by mental workload?*

Which leads to the following hypothesis  $H_2$ : *There is a significant difference in take-over time from automated to manual driving depending on various NDRT performed.*  $H_3$ : *With increasing mental workload caused by NDRT during automated driving, the ability to take-over significantly decreases.*

## 3 Methodology

### 3.1 Examined NDRT

A selection of five NDRT was evaluated by means of an online survey [40]. It was ensured that they differ in terms of their physiological load modalities. The activities to be further investigated are: Reading text (visual load), listening to radio reportage (auditory load), watching video (combination of visual and auditory load), texting (motoric and mental load) and monitoring the ride (baseline, L2 automation). To provide natural NDRT during conditional automated driving, a tablet was placed on the centre console of the vehicle. We made sure that the text is displayed in sufficient font size (about 150 words per DIN A4 page). A radio report was selected for auditory NDRT, which was a podcast for travellers.

When choosing the right content for the NDRT, watching video, movies and TV shows were excluded to

avoid that the test persons already knew them. For this reason, a scientific magazine was selected. To create the highest possible degree of authenticity in texting, the study supervisor was integrated into the experimental setting. A chat program was opened on the tablet, which enabled the subjects to communicate with the supervisor. This included chatting about their favourite food or the last holiday destination. The last activity does not offer the test person any other tasks in this setting apart from the pure monitoring of the driving. To ensure that the people perform the NDRT, check questions were asked about the content at the end of a run. To increase motivation to prioritise the NDRT, the participants were promised a higher financial reward if they answered at least half of the control questions during the NDRT correctly. Two subjects, who answered less than 40% of the primary task questions correctly during the particular NDRT, were excluded from the data analysis.

## 3.2 Metrics

### 3.2.1 Workload measurement

Since the informational stress and strain cannot be measured directly, mental workload measurements are used as suggested by [14, 15]. Subjective, psychophysiological and performance measurement approaches were used in this study and are presented below.

The subjective measure is based on the assumption that the respondents are best able to assess their mental workload themselves [41]. Subjective mental workload measurement methods are popular because of their practical advantages, e.g. the low cost, as no equipment is required and high ease of use. The National Aeronautics and Space Administration Task-Load Index (NASA-TLX) by [42], the Subjective Workload Assessment Technique (SWAT) by [43] and the Workload Profile (WP) after [44] are the most frequently used subjective methods for mental workload measurement. According to [45] the NASA TLX has a high validity, reliability and user acceptance compared to SWAT and WP, as well as a high diagnostic accuracy in dynamic environments. Furthermore, [46] show that the NASA TLX has a high sensitivity and is considered more sensitive compared to other subjective evaluation scales. Because of this, the NASA TLX is used to measure workload in this study. An increasing value correlates with an increasing load. According to a meta-analysis by [47], overstraining can occur if the overall NASA TLX score is 60 and higher; under 37 points understraining occur.

Psychophysiological measures include both the measurement of the physiological reactions of individuals to task performing and the relationship between psychological processes and their underlying physiological characteristics [48]. The physiological responses of the organism are activated autonomously and therefore

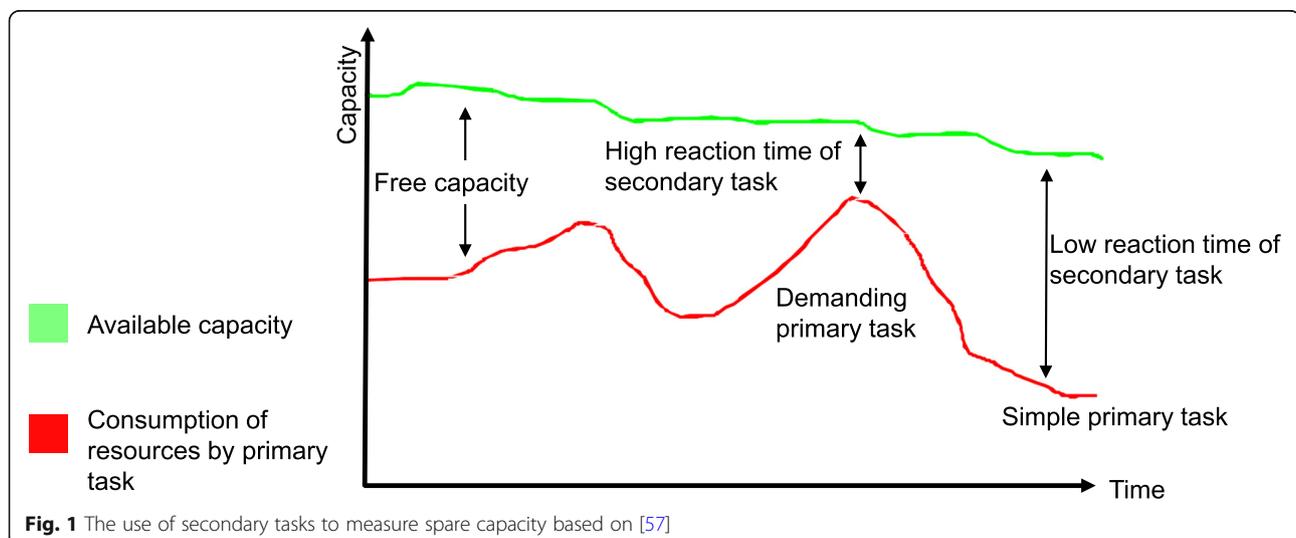
unconsciously by the peripheral nervous system. Advantages result both from the continuous measurement as well as from the small to non-existent interference with the task fulfilment [15, 49]. In addition to the advantages mentioned above, there are also limitations, since other influences such as physical stress, environmental conditions and the individual condition of the subject also affect the measurement results. An electrocardiogram (ECG) records the electrical activity of the heart over time. Relevant for the recording are the R-spikes, which describe the highest positive peak in the ECG signal. The Heart Rate Variability (HRV) is a physiological parameter for mental workload. Based on the R-R interval, heart rate variability is described over time [50]. With increasing load, the differences in R-R distances are reduced and the HRV decreases. According to [51–53] HRV decreases under both informational and physical load. The VarioPort measuring system from Becker Meditec GmbH was used to determine the psychophysiological load parameters.

Another possibility is to determine mental workload through performance measures [15]. developed a model based on the inverted U-function of optimal arousal from [54] which connects mental workload to task performance. Typical performance parameters of driving tasks are the average speed, the standard deviation of the speed or the time distance to the vehicle in front (time-to-collision). However, during automated driving and the assessment of NDRT, these driving context-related parameters cannot longer be used. To be able to measure mental workload with performance measures, it is appropriate to measure the spare capacity of mental workload. Therefore, a secondary task for the subject is added. Secondary tasks such as reaction time tests or time estimation tasks are usually found in the literature [55]. Furthermore, measurement with secondary tasks

can be divided into two paradigms [56]. With the Loading Task Paradigm, the performance of the secondary task is to be maintained, the performance loss of the primary task is thereby measured. Within the second paradigm, the Subsidiary Task Paradigm, the subject is instructed to avoid deterioration in the performance of the primary task at the expense of the secondary task.

Depending on the primary task demand, resources are required from the primary task. Due to the fact that resources are limited [17], only the remaining capacity can be used to perform the secondary task. Consequently, the performance of the secondary task varies depending on the task load of the primary task. This difference in performance of the secondary task is measured and can be compared. Figure 1 illustrates that the task load in the form of resource consumption is a fluctuating curve. The task demands are therefore interpreted as a continuum rather than a steady state (cf. [58]). If no differences in secondary task performance are measured for tasks of varied complexity, this may be caused by the subject choosing the priority of the task incorrectly and in favour of the secondary task (change from Subsidiary Task Paradigm to Loading Task Paradigm).

For the study, a Detection Response Task (DRT) according to [59] is chosen, taking the Subsidiary Task Paradigm into account. The participants in the experiment must react to a stimulus that occurs randomly every 3 to 5 s for approximately 2 min, by pressing a button. The stimulus is emitted for 1 s or until the participant returns a positive response. A valid response to a stimulus exists if the subject presses the button within 100–2500 ms after the stimulus begins. Unrealistic responses below 100 ms and responses longer than 2500 ms were not evaluated and were coded as a fault. This value is included in the calculation of the percentage hit-rate.



The visual stimulus (LED 5 mm, light colour 626 nm) was head-mounted at 12 to 13 cm to the left eye. This head-mounted variant offered the advantage that the stimulus was always in the same position in the field of vision even during head movements. The response button is contrary to [59] located in a comfortable position on the left armrest of the driver's door instead of the finger itself. This adjustment was necessary due to the design of the NDRT and an enhanced cable rupture protection.

### 3.2.2 Take-over controllability

During conditional automated driving, the vehicle user must be able to respond to a TOR from the system at any given time and take over vehicle control [60]. In this paper we will only focus on the time factor in take-over controllability. However, time is not the only consideration, the quality of take-over also has a crucial role in this context. For more information, see [40]. In this paper, the term take-over time is used to describe the minimum take-over time. This is the time difference between the start of the TOR and the minimum time value of the steering or braking intervention. A brake engagement was classified as such if the brake pedal was moved by at least 10%. For steering intervention, a change in the steering angle of at least 3° has been found to be appropriate (cf. [61]). Generally, shorter reaction times correlate with better take-over controllability.

### 3.3 Apparatus

At the time of this research, neither a production nor prototype test vehicle was available that could meet the conditional automated driving characteristics as defined by [5]. Therefore, the test trials were carried out on the static driving simulator at the Institute of Ergonomics and Human Factors at TU Darmstadt. The driving simulator consists of a fully assembled vehicle mock-up (Chevrolet Aveo, 2008) surrounded by six projection screens. Three front projection screens provide a forward and side view and another three provide a view of

the rear traffic, which the test person can see through the existing exterior and interior mirrors, see Fig. 2.

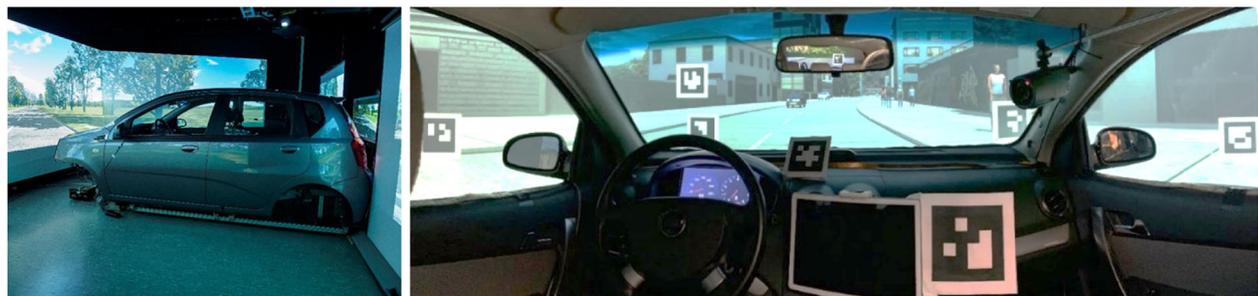
We used the Silab simulation software by WIVW GmbH for this study. A automation controller for conditional automated driving according to [5] was developed for this investigation. This provided a standardized and thus comparable test drive for each participant. During the automated drive, the driver can intervene at any time and override the automation system.

### 3.4 Driving scenario

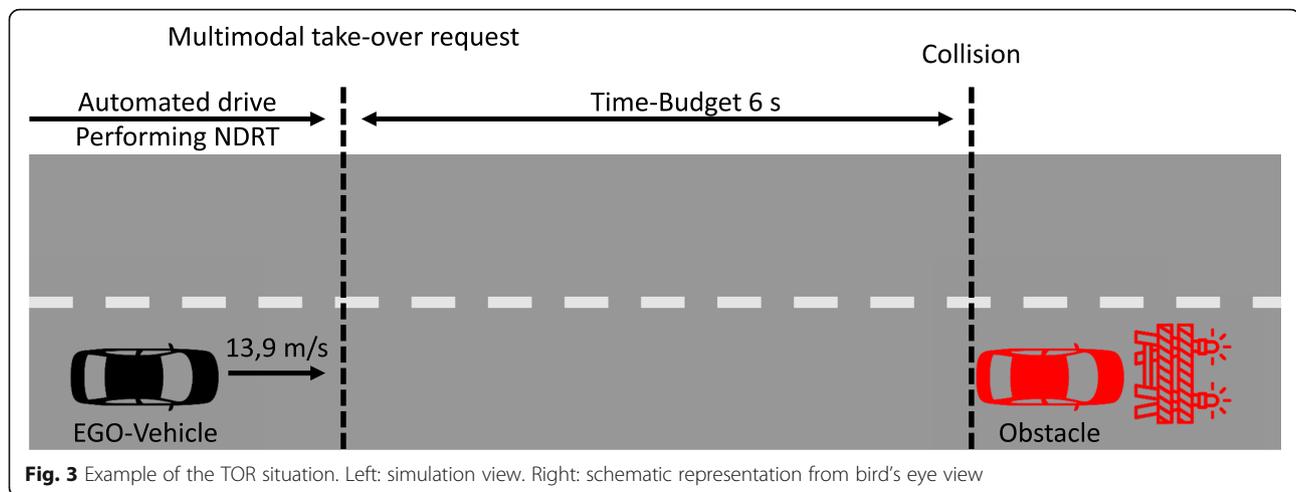
For each NDRT to be investigated, a separate urban route was designed. According to [62], a typical urban route has characteristics such as a permissible maximum speed between 30 and 50 km/h, a rather high traffic density, traffic light systems, increased number of road signs as well as turning and braking procedures. The simulated urban route has a length of approximately 19 min (9 km) for each NDRT. To ensure that the participants cannot anticipate an impending TOR, the order of the individual test sections in the route design and the traffic routing was varied for each NDRT. For all five TOR scenarios, no additional traffic was added to keep the influence factor of traffic density constant. The TOR takes place for each NDRT on a straight section of road at a speed of 13.8 m/s after passing a pre-defined waypoint. During the actual TOR, the subject must prevent an impending collision by evading or braking. After driving around the obstacle, the automation controller is reactivated in the original lane and the subject can continue with the NDRT. A schematic overview of a TOR is shown in Fig. 3. During the measurement section of the mental workload and the secondary task, the automated vehicle drove along the city route and no further incidents occurred.

### 3.5 Study design

Given the high number of variables of the constructs investigated, the decision was made to use a dependent sample in a within-subjects study design [63]. In this



**Fig. 2** Left: Exterior view of the static driving simulator at the Institute of Ergonomics & Human Factors at the TU Darmstadt; Right: View of the vehicle interior



case all subjects perform all NDRT in a permuted order. The vehicle was always driven in an automated mode and one of the five NDRTs was performed. For each NDRT to be examined, the trial run is divided into three sections: 1) psychophysiological measurement, 2) performance measurement with a secondary task and 3) TOR. After the psychophysiological measurement, the NASA TLX questionnaire was answered by the participants for the subjective workload measurement.

According to the time requirement of [64] a 7 min section for the psychophysiological measurement was chosen. Since [34] could not detect any effect on the take-over performance after a short trip (5-min) compared to a longer trip (20-min), the TOR was carried out within a 5-min section. A visual, auditory and vibrotactile TOR that had already been empirically evaluated (cf. [61]) was used in this study. A red steering wheel icon was projected onto the road using a head-up display; a warning tone was emitted through the in-car audio system and a vibration was generated by the in-seat motors. All three alert stimuli were delivered simultaneously and did not differ across the study. The secondary task is simultaneously carried out with the NDRT in a seven-minute section, see Fig. 4. Before the actual investigation began an acclimatisation, the drivers were intended to get used to the simulator and were already presented with an exemplary TOR. After the approximately five-minute training drive, reference measurements for the psychophysiological measuring and for the secondary task were carried out without performing a NDRT nor automated driving of the car.

### 3.6 Data analysis

The measured parameters are displayed in Boxplot diagrams. The significance tests were selected based on a decision tree from [65]. The significance level was set to  $\alpha = 0.05$ .

Since several NDRT were examined, an ANOVA with repeated measurements were used. If the standard deviations within the NDRT differ, a Greenhouse-Geisser correction was applied. As soon as significant differences were found, post-hoc tests were then performed to determine differences between the individual NDRTs.

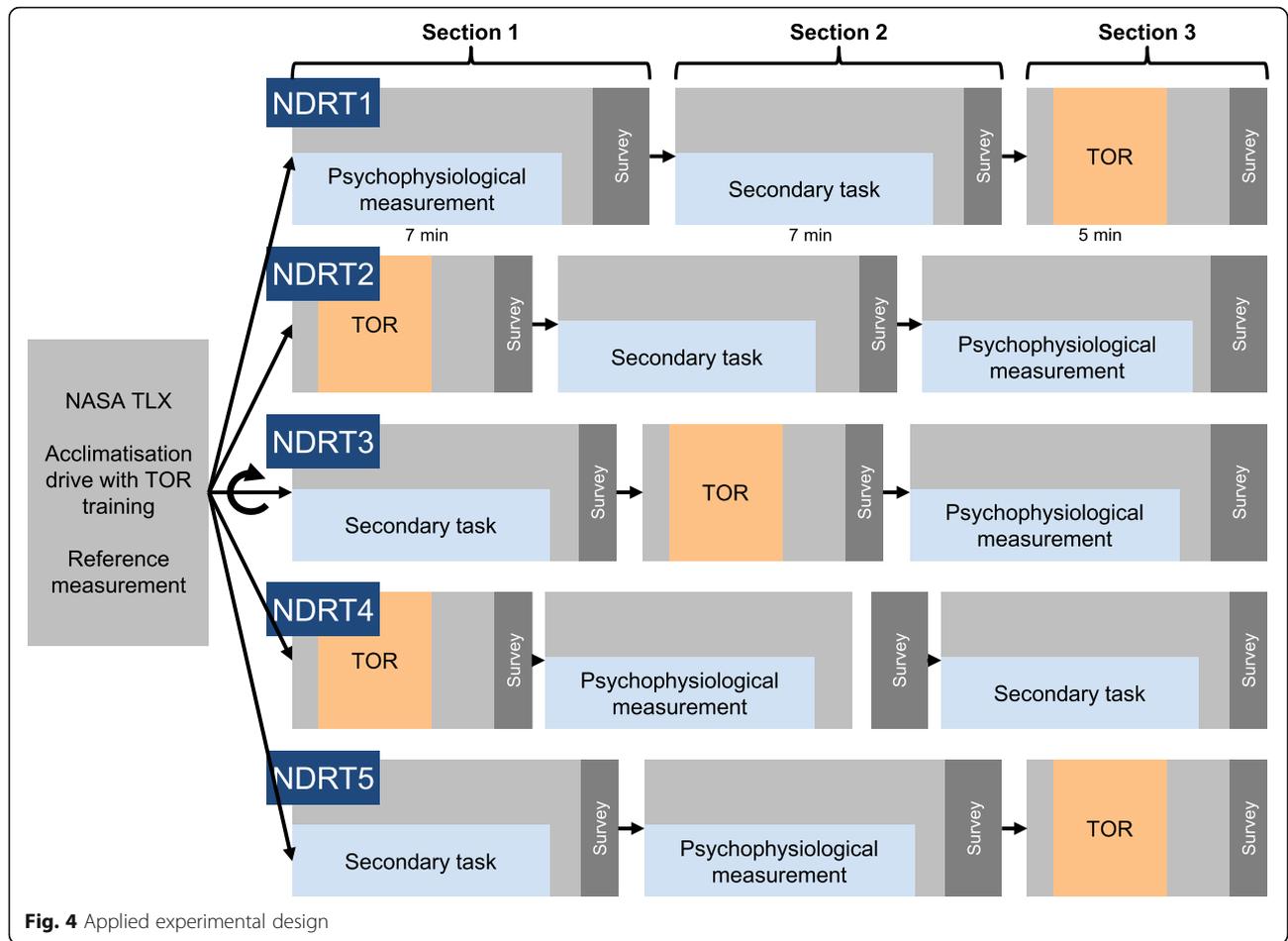
## 4 Results

### 4.1 Sample details

Sixty-two subjects could be recruited for the study. Six of them had to prematurely stop the trial due to simulator sickness and were not included in the data analysis and for two persons, one NDRT data set each had to be excluded as the persons did not achieve sufficient results on the NDRT control questions. The participants were distributed in almost equal proportions across the gender. A total of 30 male (53.6%) and 26 female (46.4%) participants were part of the study. The subjects were 19–59 years old and had an average age of 33.2 years (SD 12.0). The experiments took place in February and March of 2019.

### 4.2 Mental workload

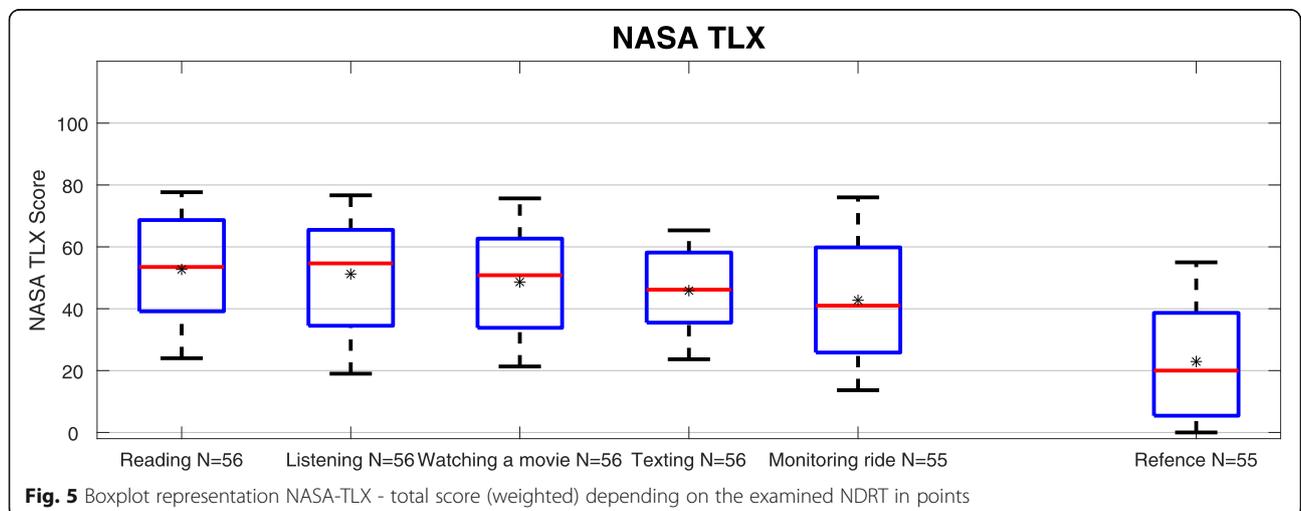
The evaluation of the subjective workload was carried out using the NASA-TLX. The weighted overall evaluation shows that the workload for the NDRT *Reading* is the highest with 52.47 (SD = 17.68 points) of 100 possible points. The detailed results are shown in Fig. 5 and listed in Table 1 (mean value and standard deviation as well as the respective post-hoc tested mean value difference). A significant higher mental workload can be recognized between the reference measurement (23.24 points, SD = 19.11) and all NDRT. The result of the multifactorial analysis of variance with repeated measurements confirms this significant difference between the tested factors [F (5, 265) = 28.67,  $p < 0.001$ ,  $f = 0.37$ ]. The workload



**Fig. 4** Applied experimental design

depending on the NDRT shows a significant difference only between the tasks *Monitoring Ride* and *Reading*. The arithmetic mean of the perceived workload decreases in the following order: Reading, Listening, *Watching a movie*, *Texting* and *Monitoring ride*.

Objective measured workload is given in this paper by the heart rate variability (HRV) parameter rMSSD. A low rMSSD value indicates a higher mental workload. In comparison to all other activities, significantly lower values can be found for *Texting* (33.64 ms, SD = 16.14



**Fig. 5** Boxplot representation NASA-TLX - total score (weighted) depending on the examined NDRT in points

**Table 1** Overview of the results NASA-TLX - total score (weighted) in points

NDRT	MW	SD	Reading	Listening	Watching a movie	Texting	Monitoring ride
Reading	52,47	17,68					
Listening	50,79	19,58	1,67				
Watching a movie	48,64	17,85	3,83	2,15			
Texting	45,72	13,88	6,74	5,06	2,91		
Monitoring ride	42,24	20,31	<b>10,22<sup>a</sup></b>	8,54	6,39	3,48	
Reference	23,24	19,11	<b>29,22<sup>a</sup></b>	<b>27,54<sup>a</sup></b>	<b>25,39<sup>a</sup></b>	<b>22,48<sup>a</sup></b>	<b>19,00<sup>a</sup></b>

<sup>a</sup>significant difference in mean value difference ( $\alpha = 0.05$ ; Bonferroni corrected)

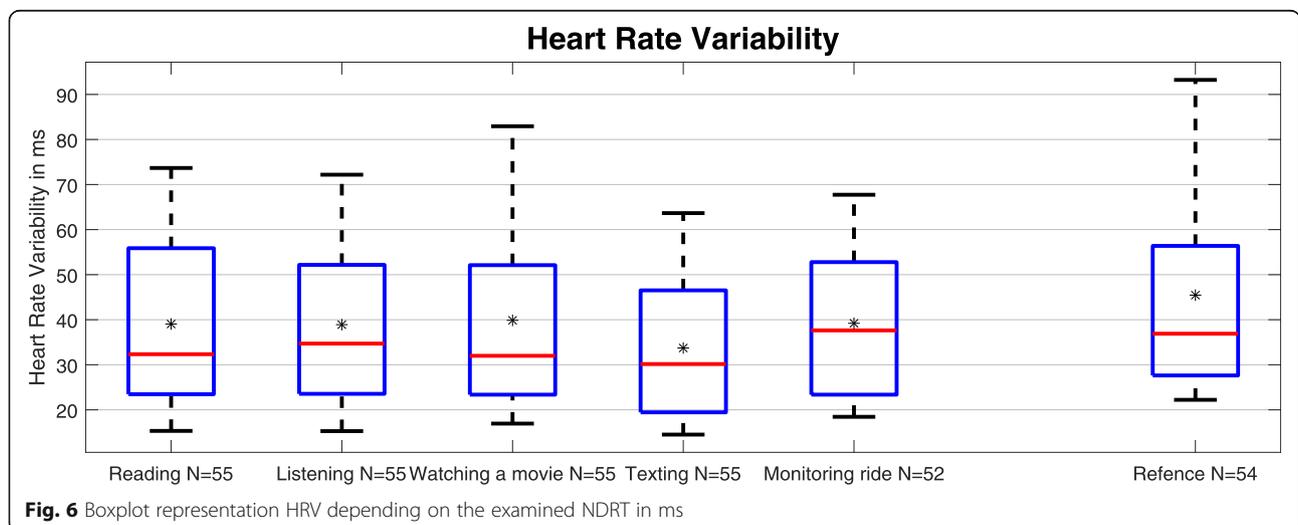
ms). The *reference measurement* (46.25 ms, SD = 22.87 ms), on the other hand, has significantly higher rMSSD values, which can be associated with lower mental workload. The distribution is shown in Fig. 6 and listed in Table 2. All other NDRT are in a similar range and do not differ significantly from each other. A Greenhouse-Geisser correction was applied because of different variances of the NDRT. This led to a low effect [Greenhouse Geisser  $F(3.76, 191.93) = 16.62, p < 0.001, f = 0.24$ ].

The results of the performance-based workload characteristics determined by the secondary task are presented in the following. As recommended by [59], the analysis focus lies on the response times, since the hit-rates did not distinguish among the examined NDRT. Longer reaction times in the secondary task correlate with a higher mental workload and a higher attention allocation of the NDRT. The longest reaction times are found in *Texting* (536.73 ms, SD = 104.80 ms, hit-rate = 94,15%) and in *Reading* (455.39 ms, SD = 113.20 ms, hit-rate = 97,26%). *Watching a movie* follows with an average reaction time of 361.59 ms (SD = 74.27 ms, hit-rate = 99,57%). *Listening* (346.92 ms, SD = 79.24 ms, hit-rate = 100%) and *Monitoring ride* (336.34 ms, SD = 61.90 ms, hit-rate = 98, 97%) both have lower reaction times. The *reference*

*measurement* showed the lowest response times with 267.74 ms (SD = 46.05 ms, hit-rate = 100%). After a Greenhouse-Geisser correction, a very large effect resulted, which depends on the individual NDRT [ $F(3.29, 164.85) = 154.13, p < 0.001, f = 1.13$ ]. The further analysis confirms that there are no significant differences between the NDRT *Listening* and *Watching a movie* and between *Listening* and *Monitoring ride*. All results are shown in detail in Fig. 7 and in Table 3.

**4.3 Take-over time**

The parameter Take-over time results from the time difference between TOR and steering or braking intervention by the participant. The value should be as low as possible to be able to claim a good take-over capability. The longest average minimum take-over time could be determined for the NDRT *Reading* (1.64 s, SD = 0.31 s, 90th percentile = 2.48 s). This is significantly longer than all other NDRT examined. *Watching a movie* (1.48 s, SD = 0.24 s, 90th percentile = 2.49 s) and *Texting* (1.49 s, SD = 0.28 s, 90th percentile = 3.42 s) are in a very similar range and the take-over time is significantly longer than with the NDRT *Listening* (1.10 s, SD = 0.25 s, 90th percentile = 1.83 s) and *Monitoring ride* (1.11 s, SD = 0.38 s, 90th percentile = 3.02 s). The latter two NDRT do not



**Table 2** Overview of the results HRV in ms

NDRT	MW	SD	Reading	Listening	Watching a movie	Texting	Monitoring ride
Reading	39,11	19,49					
Listening	39,10	18,29	0,01				
Watching a movie	39,92	21,62	-0,80	-0,81			
Texting	33,64	16,14	<b>5,47<sup>a</sup></b>	<b>5,45<sup>a</sup></b>	<b>6,27<sup>a</sup></b>		
Monitoring ride	39,20	17,06	-0,08	-0,09	0,71	<b>-5,55<sup>a</sup></b>	
Reference	46,25	22,87	<b>-7,13<sup>a</sup></b>	<b>-7,14<sup>a</sup></b>	<b>-6,33<sup>a</sup></b>	<b>-12,60<sup>a</sup></b>	<b>-7,04<sup>a</sup></b>

<sup>a</sup>significant difference in mean value difference (α = 0.05; Bonferroni corrected)

differ from each other. Considering the 90th percentile, it is evident that *Texting* leads to the longest take-over time. An adjusted analysis of variance shows a strong effect, [Greenhouse Geisser  $F(3.32, 172.86) = 54.46, p < 0.001, f = 0.59$ ]. The results are given in Fig. 8 and Table 4.

### 5 Discussion and conclusion

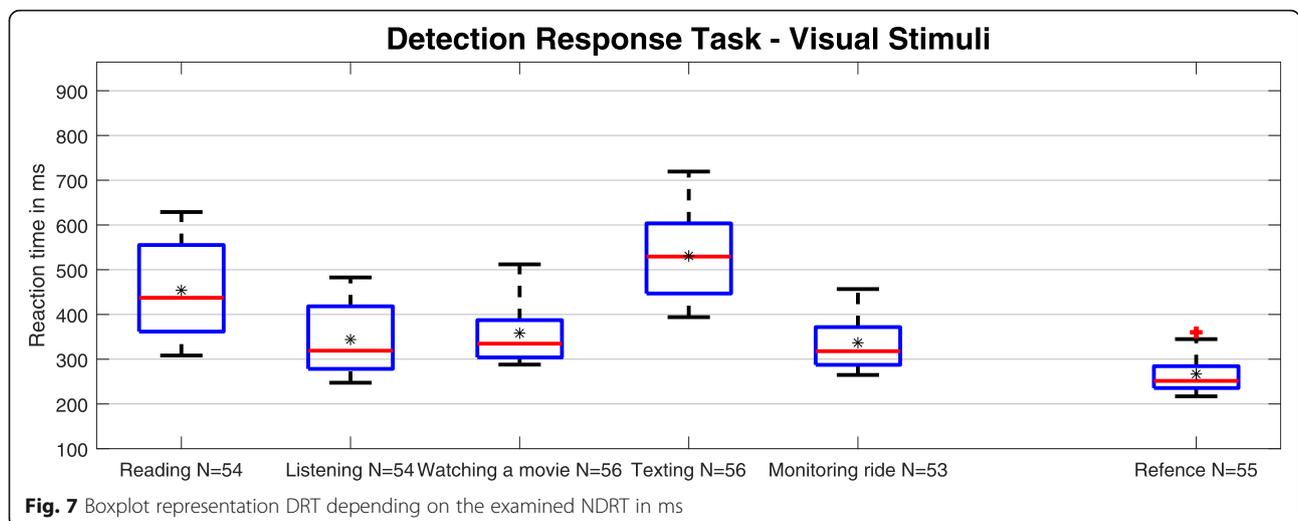
The results presented are used at this point to answer the research questions presented at the beginning.

The trimodal approach of subjective, psychophysiological, and performance-based measurement methods was used to assess mental workload. The methods used are reviewed below and the results are discussed in closing. At the beginning of the experiment a reference measurement was carried out for all mental workload characteristics to establish comparability. As this reference measurement was performed in the paused simulator, the test participants might have felt an initial excitement due to the unknown situation. As a result, there may be a bias in the subjective perception and in the psychophysiological data.

The perceived workload was measured using a NASA TLX questionnaire. Since no data were available yet on the actual performing of naturalistic

NDRT during automated drive, these results can be used as a first data basis for further research purposes. A weighting of the individual six dimensions was carried out for each NDRT by pair comparison. Due to the differentiated scores in the individual categories, the total score does not provide clear indications regarding the mental workload of each NDRT. More information can be found in [40]. Benefits of the method include easy handling. Despite the dimensional description in the questionnaire, there may have been errors in answering the questionnaire and thus a misjudgement of the respondents.

The psychophysiological data collection for mental strain measurement turned out to be less reliable due to the high variance. A clear distinction as to which NDRT are more demanding cannot be satisfactorily assessed at this point with the measurement methods used. During the actual examination of NDRT, the measurement of cardiovascular activity was carried out in such a way that the physical load was as low as possible. As the psychophysiological measurement showed, *Texting* was the most demanding compared to the other NDRTs. Since the typing also involved the motor part of the hand-arm system, it can be argued that this may have resulted in a lowered HRV. A clear distinction between physical,



**Fig. 7** Boxplot representation DRT depending on the examined NDRT in ms

**Table 3** Overview of the results DRT in ms

NDRT	MW	SD	Reading	Listening	Watching a movie	Texting	Monitoring ride
Reading	455,39	113,20					
Listening	346,92	79,24	<b>108,47<sup>a</sup></b>				
Watching a movie	361,59	74,27	<b>93,80<sup>a</sup></b>	- 14,67			
Texting	536,73	104,80	<b>-81,33<sup>a</sup></b>	<b>- 189,80<sup>a</sup></b>	<b>- 175,13<sup>a</sup></b>		
Monitoring ride	336,34	61,90	<b>119,05<sup>a</sup></b>	10,58	<b>25,25<sup>a</sup></b>	<b>200,39<sup>a</sup></b>	
Reference	267,74	46,05	<b>187,65<sup>a</sup></b>	<b>79,18<sup>a</sup></b>	<b>93,85<sup>a</sup></b>	<b>268,99<sup>a</sup></b>	<b>68,60<sup>a</sup></b>

<sup>a</sup>significant difference in mean value difference ( $\alpha = 0.05$ ; Bonferroni corrected)

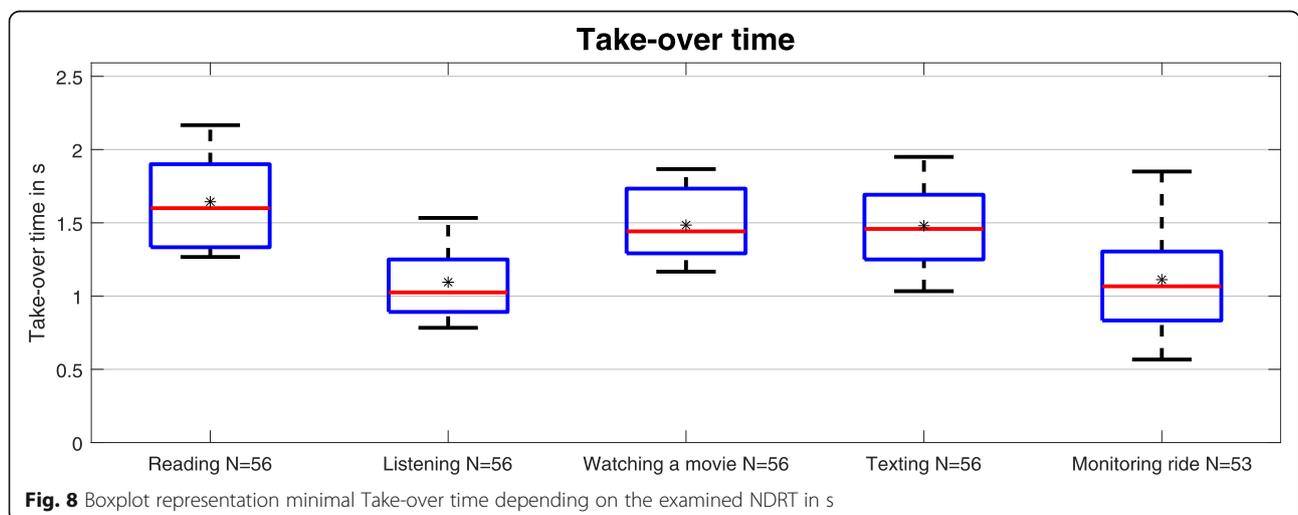
mental or emotional load is not possible when evaluating the characteristics of the electrocardiogram, so that influences of physical and emotional load on the mental workload cannot be excluded.

A disadvantage of the secondary task method is the increase in load caused by the DRT itself, since it must be considered as an independent load [66]. However, even if the informational processing requirement of the stimulus-response time test can be regarded as minimal and can be learned quickly, it cannot be excluded that the DRT may bias the simultaneous measurement of psychophysiological data. But even if this is the case, this is not relevant, since it is not the absolute values that are considered but rather the relative comparison between the NDRT. The participants were fast to understand the function of the DRT. According to the Subsidiary Task Paradigm, the performance drop should only occur in the secondary task. The DRT proved to be a very sensitive measuring tool, since it is very well able to recognize even small differences, cf. [67]. For example, in this study, significant higher mental workload expressed by a longer reaction time could already be observed at the *Monitoring ride* in comparison to the reference measurement. Furthermore, the DRT reaction-times revealed significant measurement differences and small variances

compared to the psychophysiological measurements. On the other hand, no significant differences were found at the DRT hit-rate.

After discussing the methods in detail, we will summarize these below. This study could prove that the mental workload differs depending on the NDRT while conditional automation driving. For the aforementioned reasons, the hypothesis  $H_1$  cannot be refuted. The subjective workload perception for each NDRT investigated differs significantly from the reference measurement taken during vehicle standstill without NDRT. It was found that *Reading* was perceived as the most demanding NDRT. However, all examined NDRT showed a high variance, so that a clear distinction is not possible. In addition, the single dimensional analysis showed that the test participants enjoyed *Texting* in particular, as they indicated a lower frustration level, which can explain the comparatively subjectively low perceived feeling of workload.

The psychophysiological parameters also show a high variance among each other and can additionally react sensitively to emotional and physical stress. Cardiovascular activity in the form of HRV has been identified in various literature sources as a mental workload indicator. Significant differences in NDRT were also found in



**Table 4** Overview of the results minimal Take-over time in s

NDRT	MW	SD	Reading	Listening	Watching a movie	Texting
Reading	1,64	0,31				
Listening	1,10	0,25	<b>0,54<sup>a</sup></b>			
Watching a movie	1,48	0,24	<b>0,16<sup>a</sup></b>	<b>-0,38<sup>a</sup></b>		
Texting	1,49	0,28	<b>0,15<sup>a</sup></b>	<b>-0,39<sup>a</sup></b>	-0,01	
Monitoring ride	1,11	0,38	<b>0,53<sup>a</sup></b>	-0,00	<b>0,37<sup>a</sup></b>	<b>0,38<sup>a</sup></b>

<sup>a</sup>significant difference in mean value difference ( $\alpha = 0.05$ ; Bonferroni corrected)

this study. *Texting* shows a significantly higher load despite the high variance of the measured values.

The results of performance-based workload measurement show similar results to those of psychophysiological measurements. *Texting* is also the most demanding activity. *Reading* also proves to be more demanding than *Watching a movie*, *Listening* or *Monitoring ride*.

However, the results of the different measurement methods are not all consistent in their own aspects. Large differences were found in the self-assessment instrument NASA-TLX, because *Texting* was found to be less mentally demanding. This finding is of great interest, because in literature the workload is often only represented by the NASA TLX. This leads to possible falsifications, as objectively measured activities with a high workload are perceived as comparatively less demanding due to the actual joy of use (expressed by a low level of frustration). A sensitive tool to determine the mental workload of a NDRT during automated drive is the performance-based measurement by means of competing secondary tasks in form of the DRT. *Reading* and *Texting* were consistently identified by psychophysiological and performance-based measurement methods as the most demanding NDRT. The subjective perception confirms this only for the NDRT *Reading*. *Listening*, *Watching a movie* and *Monitoring ride* showed no significant differences at the psychophysiological and performance-based parameters.

An essential question in the investigation of conditional automated driving systems is whether the user of the vehicle can quickly take back control of the vehicle in case of a TOR.

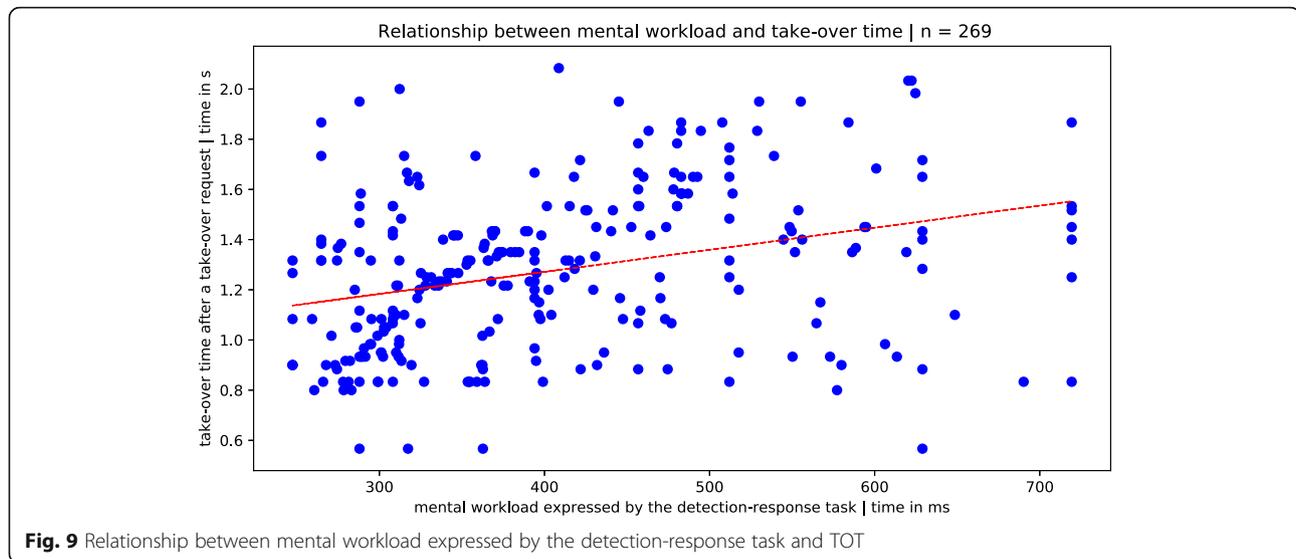
As described in the theoretical part, TOT depends on many factors. The influence of the design of the TOR was empirically examined in advance of this study and the best version was chosen [61]. Through a training drive before the actual trial, the participants were already confronted with a TOR, allowing them to gain a sufficient knowledge of the system. A special aspect of this study is the consideration of an urban scenario. There have been studies (such as in [29]) that showed that a more complex traffic situation is resulting in longer TOT. The purpose of this study was to simulate a complex traffic situation in the

urban scenario, so that a worst-case situation could be investigated. Therefore, even longer TOT can be ruled out due to the traffic scenario. The TOR was carried out on a straight to be able to differentiate between measured steering angle changes initiated by the test person and the target steering angle of the automation controller. For following experiments, the investigation area should be similarly structured as in this study. A time budget of 6 s ensured that there was no collision with the obstacle and should be shortened in future so that more significant results on take-over can be achieved. In total, the test participants experienced six TOR over the entire study. No learning-effects in dependence of the number of TOR experienced could be determined.

Significant differences were found in the minimum take-over time parameter depending on the performed NDRT. Therefore hypothesis H<sub>2</sub> cannot be refuted either. The mean of the minimum take-over time for this study are between 1.10 s (*Listening*) and 1.64 s (*Reading*) and therefore in a shorter range compared to the presented literature.

Moreover, the results of this work differ from studies that have also examined multiple NDRT. In [25] no differences were found between the take-over times of two NDRT [39]. also examined the influence of naturalistic NDRTs (writing e-mails, reading messages and watching videos) on take-over performance. In their study, no significant differences in take-over times were found in relation to the NDRT examined. A possible explanation for this might be the fact that the NDRT were explicitly the focus in this experiment and that the participants were questioned about the content of the NDRT. This meant that they were even more involved in the implementation of the NDRT.

In a more detailed analysis, the relationship between the construct mental workload in relation to the take-over time was identified. A regression analysis reveals that the take-over time increases significantly with increasing mental workload [ $F(1, 268) = 30.74, p < 0.001, R^2 = 0.103$ ], see Fig. 9. Consequently, NDRT with high mental workload lead to longer reaction times. Hence the hypothesis H3 cannot be refuted.



**Fig. 9** Relationship between mental workload expressed by the detection-response task and TOT

In contrast to the study of [68], a significant correlation between mental workload in the form of DRT and the ability to react in a critical situation was found in this study. To ensure a better TOT, it can be concluded from this study that individuals should not have a high mental workload. The mental workload can be influenced significantly by the task difficulty. A mental workload that is too low can lead to insufficient demand and thus to monotony-induced fatigue (cf. [69, 70]), which should also be avoided. Due to the mentioned conflicts between the execution of NDRT and a short take-over time, the question arises if the automation level 3 is a desirable automation approach or if NDRT should only be allowed from level 4 onwards, where the individuals do not have to intervene in the driving process anymore.

### 5.1 Limitations

Every study has certain limitations that must be considered when interpreting the results. Since no vehicle with the appropriate level of automation was available, the research had to be carried out in a driving simulator. However, this has the consequence that the transferability to real road traffic may be limited. For example, the participants may have behaved differently in the simulator due to a potentially increased feeling of safety. In this case, it could be possible that individuals are less likely to engage with the NDRT and that in field experiments the results could differ from each other. In order to verify this, a questionnaire was handed out at the end of the experiment to evaluate the test setup. The average answer to the question “In real road traffic I would have behaved differently in an automated vehicle” is 3.49 (SD = 1,20; 1 = Not applicable at all; 5 = Fully applicable). This confirms the hypothesis recently made. It was

reported that they did not trust the automated system and therefore continued to monitor the automated ride. Even though the measured absolute values could differ from reality, the identified relative differences between the NDRT can be transferred to real world conditions [71]. Due to the number of variables and the sample size, a Within-Subjects Study Design was used. Here, especially learning effects must be considered. To counteract this, the course was designed differently for each NDRT. The duration of the experiment was between 3.5 and 4 h per person, depending on how much time the subjects needed to answer the questionnaires. Overall, there were sufficient breaks between the sections for food intake and recovery. However, some participants in the experiment stated that they felt that the experiment was taking too long. To ensure good transferability to real-world practice, the focus of this research is on naturalistic NDRT. However, the drawback of naturalistic NDRT as opposed to standardised (e.g. n-Back or SuRT) is that it is less comparable with other studies.

### 6 Conclusion

In a simulator study, the effects of NDRT in conditional automated driving were investigated. In contrast to partially automated vehicles, where the driver must monitor the driving situation continuously, the users can turn away from monitoring the road. However, the vehicle user is still considered as a fallback level and must therefore be ready to take over the vehicle quickly and safely. But the parallel execution of NDRT and a short take-over time in critical situations conflict with each other. Therefore, the gain in comfort, handling NDRT's, results in reduced take-over controllability. To ensure a better

take-over controllability, it can be concluded from this work that vehicle users should not be exposed to high mental workload. A lack of mental workload can lead to understress and thus to monotonous fatigue, which should also be avoided. This can be achieved, for example, through gamification and the targeted use of NDRT [69, 70].

Due to the conflicts mentioned between the execution of NDRT and a short take-over time, the question arises in summary, whether the automation level of conditional automated driving is an approach to be striven for or whether NDRT should only be permitted from even higher level of automation, where human intervention in the driving process is no longer required.

#### Acknowledgements

The Ethics Committee of the Technical University of Darmstadt has granted the ethical approval for this study. The application identification code is EK 55/2018.

We acknowledge support by the German Research Foundation and the Open Access Publishing Fund of *Technical University of Darmstadt*.

#### Authors' contributions

ALM was leading the research team on "Effects of Non-driving Related Tasks on mental workload and take-over times during conditional automated driving". He wrote the article by himself with the data provided by following authors. NFE wrote her master thesis in the research team about mental workload and analyzed mentioned data. She also supervised parts of the simulator study. RH wrote his bachelor thesis in the research team about the detection response task and analyzed mentioned data. He also supervised parts of the simulator study. LZ wrote his bachelor thesis in the research team about take-over controllability and analyzed mentioned data. He also supervised parts of the simulator study. AB is the deputy director of the institute and has proofread the article and advised on changes. She is now the corresponding author. All authors read and approved the final manuscript.

#### Funding

This work is a result of the research project @CITY – Automated Cars and Intelligent Traffic in the City. The project is supported by the Federal Ministry for Economic Affairs and Energy (BMWi), based on a decision taken by the German Bundestag. The authors are fully responsible for the content of this study and publication. We further acknowledge support by the German Research Foundation and the Open Access Publishing Fund of *Technical University of Darmstadt*.

#### Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

#### Competing interests

The authors declare that they have no competing interests.

Received: 29 October 2020 Accepted: 4 February 2021

Published online: 20 February 2021

#### References

- Lemmer, K. (2014). Einführungsvortrag "Autonomes Fahren": Deutsches Zentrum für Luft- und Raumfahrt e.V.
- Bartels, A., To T-B, Karrenberg, S., & Weiser, A. (2011). Hochautomatisches Fahren auf der Autobahn. *ATZ Automobiltech Z*, 113(9), 652–657.
- Carsten, O., Lai, F. C. H., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in Semiautomated driving: Does it matter what aspects are automated? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 54(5), 747–761.
- Biever, W., Angell, L., & Seaman, S. (2019). Automated Driving System Collisions: Early Lessons. *Human Factors: The Journal of the Human Factors and Ergonomics Society Special Issue on In-Vehicle Automation*: 1-11.
- SAE. J3016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles: Society of Automotive Engineers; 2018.
- Naujoks, F., Befelein, D., Wiedemann, K., & Neukum, A. (2018). A Review of Non-driving-related Tasks Used in Studies on Automated Driving. In N. A. Stanton (Ed.), *Advances in human aspects of transportation: Proceedings of the AHFE 2017 International Conference on Human Factors in Transportation, Los Angeles, California, USA*, (pp. 525–537). Cham: Springer.
- Wandtner, B., Schömig, N., & Schmidt, G. (2018). Effects of non-driving related task modalities on takeover performance in highly automated driving. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 60(6), 870–881.
- Stanton, N. A., & Young, M. S. (2000). A proposed psychological model of driving automation. *Theor Issues Ergon Sci*, 1(4), 315–331.
- Winter, J. d., Happee, R., Martens, M. H., & Stanton, N. A. (2014). Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transport Res F: Traffic Psychol Behav*, 27(Part B), 196–217.
- Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and Trust in Automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making*, 2(2), 140–160.
- Sarter, N. B., & Woods, D. D. (1991). Situation awareness: A critical but ill-defined phenomenon. *Int J Aviat Psychol*, 1(1), 45–57.
- Landau, K. (2005). *LexAB – Kleines Lexikon arbeitswissenschaftlicher Begriffe*. Stuttgart: Ergonomia Verlag.
- Luczak, H. (1975). *Untersuchungen informatorischer Belastung und Beanspruchung des Menschen*. Düsseldorf: VDI-Verlag.
- O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload Assessment Methodology. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance*, (2nd ed., pp. 1–49). Oxford: John Wiley & Sons.
- DeWaard D. The measurement of Drivers' mental workload. Dissertation, Psychologische, Pädagogische en Sociologische Wetenschappen, Universiteit Groningen 1996.
- Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, 58(1), 1–17.
- Wickens, C. D. (2002). Multiple resources and performance prediction. *Theor Issues Ergon Sci*, 3(2), 159–177.
- Ribback S. Psychophysiologische Untersuchung mentaler Beanspruchung in simulierten Mensch-Maschine-Interaktionen. Dissertation, Lehrstuhl für Arbeits-, Betriebs- und Organisationspsychologie, Universität Potsdam 2003.
- Packebusch, L. (2003). Psychische Belastung und Beanspruchung–Normung für die Praxis. *Wirtschaftspsychologie aktuell*, 3(4), 32–36.
- Zeeb, K. (2016). Der Einfluss fahrfremder Tätigkeiten auf die Fahrerübernahme während des hochautomatisierten Fahrens. Dissertation, Insitut für experimentelle Psychologie, Heinrich-Heine-Universität.
- de Winter, J., Stanton, N. A., Price, J. S., & Mistry, H. (2016). The effects of driving with different levels of unreliable automation on self-reported workload and secondary task performance. *Int J Veh Des*, 70(4), 297–324.
- Eriksson, A., & Stanton, N. A. (2017). Takeover time in highly automated vehicles: Noncritical transitions to and from manual control. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 59(4), 689–705.
- Petermann-Stock, I., Hackenberg, L., Muhr, T., & Mergl, C. (2013). Wie lange braucht der Fahrer?: Eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt. In TÜV SÜD (Ed.), *Wie lange braucht der Fahrer?: Eine Analyse zu Übernahmezeiten aus verschiedenen Nebentätigkeiten während einer hochautomatisierten Staufahrt*, (pp. 1–26).
- Zhang, B., de Winter, J., Varotto, S., Happee, R., & Martens, M. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transport Res F: Traffic Psychol Behav*, 64, 285–307.
- Vogelpohl, T., Vollrath, M., Kühn, M., Hummel, T., & Gehlert, T. (2016) Forschungsbericht Nr. *Übergabe von hochautomatisiertem Fahren zu manueller Steuerung*, (p. 39). Berlin: Gesamtverband der Deutschen Versicherungswirtschaft e. V.
- Walch, M., Mühl, K., Kraus, J., Stoll, T., Baumann, M., & Weber, M. (2017). From Car-Driver-Handovers to Cooperative Interfaces: Visions for Driver–Vehicle Interaction in Automated Driving. In G. Meixner, & C. Müller (Eds.),

- Automotive User Interfaces: Creating Interactive Experiences in the Car*, (pp. 273–294). Basel: Springer International Publishing.
27. Damböck, D., Farid, M., Tönert, L., & Bengler, K. (2012). *Übernahmezeiten beim hochautomatisierten Fahren*. München: Tagung Fahrerassistenz.
  28. Radlmayr, J., & Bengler, K. (2015). *FAT-Schriftenreihe. Literaturanalyse und Methodenauswahl zur Gestaltung von Systemen zum hochautomatisierten Fahren: Literature Survey and Description of Methods for the Development of Highly Automated Driving*, (p. 276). Berlin: FAT - Forschungsvereinigung Automobiltechnik e.V.
  29. Radlmayr, J., Gold, C., Lorenz, L., Farid, M., & Bengler, K. (2014). How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 2063–2067.
  30. Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 58(4), 642–652.
  31. Shen, S., & Neyens, D. M. (2014). Assessing drivers' performance when automated driver support systems fail with different levels of automation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 2068–2072.
  32. Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J Exp Psychol*, 55(4), 352–358.
  33. ISO/TS 14198:2019-04 (2019). *Straßenfahrzeuge - Ergonomische Aspekte von Fahrerinformations- und Assistenzsystemen - Kalibrierungsaufgaben für Methoden, welche auf Fahrerfragen zugreifen, um fahrzeuginterne Systeme zu verwenden*. Berlin: Beuth.
  34. Feldhütter, A., Gold, C., Schnieder, S., & Bengler, K. (2017). How the Duration of Automated Driving Influences Take-Over Performance and Gaze Behavior. In C. Schlick, S. Duckwitz, F. Flemisch, et al. (Eds.), *Advances in Ergonomic Design of Systems, Products and Processes*, (pp. 309–318). Berlin, Heidelberg: Springer Berlin Heidelberg.
  35. Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). Take over!: How long does it take to get the driver back into the loop? *Proceedings of the Human Factors Society Annual Meeting*, 57(1), 1938–1942.
  36. Lorenz, L., Kerschbaum, P., & Schumann, J. (2014). Designing take over scenarios for automated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 1681–1685.
  37. Körber, M., Gold, C., Lechner, D., & Bengler, K. (2016). The influence of age on the take-over of vehicle control in highly automated driving. *Transport Res F: Traffic Psychol Behav*, 39, 19–32.
  38. Naujoks, F., Mai, C., & Neukum, A. (2014). The Effect of Urgency of Take-Over Requests During Highly Automated Driving Under Distraction Conditions. In R.-L. Jang, & T. Ahram (Eds.), *Advances in Physical Ergonomics and Human Factors: Part II: 5th International Conference on Applied Human Factors and Ergonomics*, (pp. 1–8). Louisville, Ky: AHFE Conference.
  39. Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accid Anal Prev*, 92, 230–239.
  40. Müller, AL. (2020). Auswirkungen von natürlichen fahrfremden Tätigkeiten bei hochautomatisierter Fahrt. Dissertation, Institut für Arbeitswissenschaft, Technische Universität Darmstadt.
  41. Schwalm, M. (2009). Pupillometrie als Methode zur Erfassung mentaler Beanspruchungen im automatisierten Kontext. Dissertation, Philosophische Fakultät, Universität des Saarlandes 2009.
  42. Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Adv Psychol*, 52, 139–183.
  43. Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In N. Meshkati, & P. A. Hancock (Eds.), *Human Mental Workload*, (pp. 185–218). Amsterdam: Elsevier Science Publishers B.V. (North-Holland).
  44. Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, 39(3), 358–381.
  45. Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Appl Psychol*, 53(1), 61–86.
  46. Estes, S. (2015). The workload curve: Subjective mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(7), 1174–1187.
  47. Grier, R. A. (2015). How high is high? A meta-analysis of NASA-TLX global workload scores. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 59(1), 1727–1731.
  48. Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychol*, 53(1), 61–97.
  49. Manzey, D. (1998). Psychophysiologie mentaler Beanspruchung. In F. Rösler (Ed.), *Ergebnisse und Anwendungen der Psychophysiologie*, (1st ed., pp. 799–864). Göttingen: Hogrefe.
  50. Miller, S. (2001). *Workload measures*. Literature review. Iowa City.
  51. Fu, R., Guo, Y., Yang, C., et al. (2011). *Research on heart rate and eye movement as indicators of drivers' mental workload*. Washington DC: Transportation Research Board 3rd International Conference on Road Safety and Simulation.
  52. Mulder, G., & Mulder-Hajonides, W. R. (1973). Mental load and the measurement of heart rate variability. *Ergonomics*, 16(1), 69–83.
  53. Sammito, S., & Böckelmann, I. (2015). Analyse der Herzfrequenzvariabilität. *Herz*, 40(1), 76–84.
  54. Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *J Comp Neurol Psychol*, 18(5), 459–482.
  55. Gunning, D. (1978). Time estimation as a technique to measure workload. *Proceedings of the Human Factors Society Annual Meeting*, 22(1), 41–45.
  56. Schlick, C., Bruder, R., & Luczak, H. (2018). *Arbeitswissenschaft*, (4th ed., ). Berlin: Springer Vieweg.
  57. Farmer, E., & Brownson, A. (2003). *Review of workload measurement, analysis and interpretation methods: European Organisation for the safety of air navigation; CARE-Integra-TRS-130-02-WP2*.
  58. Laurig, W. (1992). *Grundzüge der Ergonomie*, (4th ed., ). Beuth: Berlin, Köln.
  59. ISO 17488:2016-10 (2016). *Straßenfahrzeuge - Fahrerinformationen und Assistenzsysteme - Erkennungsreaktionsaufgabe (DRT) für den Zugriff beabsichtigter Effekte von kognitiver Belastungen während der Fahrt*. Berlin: Beuth.
  60. Gasser, T. M., Arzt, C., Ayoubi, M., et al. (2012). BAST-Bericht). *Ergebnisse der Projektgruppe Automatisierung: Rechtsfolgen zunehmender Fahrzeugautomatisierung*, (p. F83).
  61. Müller AL, Ogrizek M, Bier LR, Abendroth B (2018). *Design concept for a tactile and visual take-over request in a conditional automated vehicle during non-driving-related tasks*. Fort A. and Jallais C. (Eds.). Proceedings of the 6th Driver Distraction and Inattention conference, Gothenburg, Sweden 15–17. (online).
  62. Zöller, I. M., Diederich, C., Abendroth, B., & Bruder, R. (2013). Fahrsimulatorvalidität - Systematisierung und quantitative Analyse bisheriger Forschungen. *Zeitschrift für Arbeitswissenschaft*, 67(4), 197–206.
  63. Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *J Econ Behav Organ*, 81(1), 1–8.
  64. Malik, M. (1996). Heart rate variability. *Eur Heart J*, 17(3), 354–381.
  65. Blankenberger, S., & Vorberg, D. (1998). *Die Auswahl statistischer Tests und Maße. Flussdiagramm Martin-Luther-Universität Halle-Wittenberg; Technischen Universität Braunschweig*.
  66. Stojmenova, K., & Sodnik, J. (2018). Detection-response task-uses and limitations. *Sensors*, 18, 1–17.
  67. Harbluk JL, Burns PC, Tam J, Glazduri V, editors. *Detection Response Tasks: Using Remote, Headmounted and Tactile Signals to Assess Cognitive Demand While Driving*; 2018.
  68. Mantzke, O., & Keinath, A. (2015). Relating the detection response task to critical events – Consequences of high cognitive workload to brake reaction times. *Production Manufacturing*, 3, 2381–2386.
  69. Neubauer, C., Matthews, G., & Saxby, D. (2012). The effects of cell phone use and automation on driver performance and subjective state in simulated driving. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 56(1), 1987–1991.
  70. Bier, LR. (2019). *Gamification zur Vorbeugung monotoniebedingter Müdigkeit bei der Fahrzeugführung-im Vergleich zur Fahrer-Beifahrerinteraktion*. Dissertation, Institut für Arbeitswissenschaft, Technische Universität Darmstadt.
  71. Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accid Anal Prev*, 34(5), 589–600.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.