Visual Distraction Effects between In-Vehicle Tasks with a Smartphone and a Motorcycle Helmet-Mounted Head-Up Display

Hilkka Grahn†
Faculty of Information Technology
University of Jyväskylä
Jyväskylä Finland
hilkka.grahn@jyu.fi

Tuomo Kujala
Faculty of Information Technology
University of Jyväskylä
Jyväskylä Finland
tuomo.kujala@jyu.fi

ABSTRACT

Besides motorists, also motorcyclists need safer user interfaces to interact with useful applications on the road. In this paper, distraction effects of in-vehicle tasks conducted with a head-up display (HUD) for motorcyclists were compared to smartphone tasks with 24 participants in a driving simulator.

Compared to the smartphone tasks, the head-up display tasks decreased the percentage of inappropriately long glances by 45 percent. The head-up display tasks were also experienced as less demanding than the smartphone tasks. Additionally, the use of head-up display for navigation did not lead to gaze concentration effects compared to baseline driving.

The head-up display is concluded to be a safer option for the tested tasks for motorcyclists than a smartphone. Based on earlier research, we assume that the use of peripheral vision allowed drivers to better maintain situational awareness during the head-up display tasks compared to the head-down smartphone tasks. In addition, the easy-to-learn haptic design of the head-up display handlebar controller could be used without vision.

CCS CONCEPTS

• Human-centered computing~User studies • Human-centered computing~Laboratory experiments • Human-centered computing~Touch screens • Human-centered computing~Haptic devices • Human-centered computing~Empirical studies in HCI

KEYWORDS

Driver distraction, visual demand, visual occlusion, occlusion distance, head-up display, head-down display, head-mounted display.

Mindtrek 2018, October 10–11, 2018, Tampere, Finland

ACM ISBN 978-1-4503-6589-5/18/10…$15.00
https://doi.org/10.1145/3275116.3275134

1 Introduction

Driver distraction and especially visual distraction has been extensively studied in recent years. Several studies have shown that visually demanding in-vehicle tasks cause visual distraction and therefore are associated with high risk of safety-critical incidents, such as near crashes and crashes (e.g., [6, 12, 26, 33]).

Existing research in this field has been limited to study visual distraction effects of in-car tasks and user interfaces (UIs) for car drivers (e.g., [1, 13, 22]). However, not only car drivers but also motorcycle drivers need, for instance, navigating aids while driving. Therefore, this study focuses on motorcycle drivers and the visual distraction associated with in-vehicle devices that are used by motorcyclists.

Compared to driving a car, driving a motorcycle is even more complex task that requires great motor skills and coordination [24]. Also, motorcyclists are one of the most vulnerable road user group. For example, in 2000, they made up less than one percent of the road traffic in the UK but suffered 14 percent of deaths and serious injuries [5].

Truong, De Gruyter and Nguyen [32] found out that motorcyclists use smartphones while driving to call, text and find information. Also, Phommachanh, Ichikawa, Nakahara, Mayxay and Kimura [27] reported that motorcyclists dial, receive calls and send text messages while operating a motorcycle.

When we add the reported phone usage to the complexity of motorcycle driving, while knowing that motorcyclists are in great risk in general in traffic, the consequences of distraction can be serious. Because of this, it is important to study the visual distraction potential of in-vehicle devices that are designed for motorcyclists and to design better user interfaces also for them to access safer the services they need on the road. It is also important that motorcyclists experience the user interfaces – designed for them – easy to use, in order to make them prefer the safer UIs over smartphones.

In this paper, we study tasks conducted with a novel motorcycle helmet-mounted head-up display (HUD) manufactured by Nuviz Inc. (https://www.ridenuviz.com/) and compare those to similar tasks with identical goals conducted with a Samsung Galaxy A3 smartphone. A novel distraction testing method introduced by Kujala and Mäkelä [17] – that categorizes in-vehicle glances to appropriate or inappropriate glances dependent on the visual demand of the road point – was used to assess the visual distraction potential of the tasks. According to several studies (see Related work), HUDs can be less distracting for drivers than head-down displays (HDD). However, HUD could cause attention capture and gaze concentration effects [34]. With this in mind, we measure also the horizontal gaze activity of the drivers during the HUD navigation tasks.

Accordingly, the three research questions for the study were:

RQ 1: Are there significant differences in the visual distraction potential between in-vehicle tasks conducted with the Nuviz head-up display and with the Android smartphone?

RQ 2: Are there significant differences in experienced workload between the Nuviz head-up display and the Android smartphone tasks?

RQ 3: Are there gaze concentration effects in the Nuviz head-up display tasks compared to baseline driving?

2 Related work

2.1 Head-up displays, head-down displays and head-mounted displays

Head-up displays for in-vehicle use have been studied previously as well but this is the first study that compares distraction effects between a HUD designed for motorcyclists and a smartphone. Head-up displays may have significant potential for reducing visual distraction by in-vehicle tasks compared to head-down displays (HDD). For example, Weinberg, Harsham and Medenica [36] compared three systems – head-down display, head-up display and auditory display – for presenting textual lists. They found out that the number of in-vehicle glances was doubled when using HDD compared to HUD. In their study, the HUD was operated with a steering wheel-mounted controller.

Villalobos-Zúñiga, Kujala and Oulasvirta [35] studied a text entry method that consisted of a physical 3x4 keypad in the steering wheel and a HUD and compared it to a touchscreen QWERTY keyboard in the center console. The results showed that the physical keypad and HUD combination allowed drivers to maintain more visual attention on the road (up to 64 %). There were also less lane deviations when using the HUD combination compared to a touchscreen keyboard.

Smith, Gabbard and Conley [29] also compared HDD and HUD. During the experiments, participants were required to conduct visual search tasks while driving. They noticed that performing tasks with HUD caused less severe decrements in driving performance than with HDD. Interestingly, they also found out that HUD affected more negatively the glance patterns on the NHTSA metrics [25] than HDD. Since there were no significant decrements on driving performance while using HUD, they concluded that people could use different visual search methods with HUDs than with HDDs. Therefore, the NHTSA guidelines [25] may not be the best practice to assess the visual distraction of in-vehicle HUDs.

Lauber, Böttcher and Butz [21] compared HUDs and head-mounted displays (HMD). HMDs have the same features as HUDs but because they are head-mounted, the information showed is always available regardless of head position. Lauber et al. [21] concluded that their study could not show any significant differences in driving performance between the tested interaction techniques. In this study, the Nuviz HUD is helmet-mounted, and thus the tested HUD is also an HMD.

Smith, Streeter, Burnett and Gabbard [30] point out that HUD interfaces should be carefully designed. HUD tasks that do not support resumability may cause even greater problems than traditional head-down displays. In general, it is wise to remember that HUDs and HMDs can also be distracting and impair driving performance [9, 10]. Even if HUDs and HMDs may allow the driver to keep peripheral vision available to serve the goals of the driving task, these may cause negative gaze concentration effects [34] compared to driving without any secondary tasks.

2.2 Visual distraction - operationalization and measurement

For instance, Foley, Young, Angell and Domeyer [7] (p. 63) have defined visual distraction as follows: "*Visual distraction is any glance that competes with activities necessary for safe driving*". Among others, Klauer, Dingus, Neale, Sudweeks and Ramsey [12] have reported in their naturalistic driving study that over two seconds eyes-off-road durations – that is, visual distraction – were associated with near-crash and crash risk.

In this study, we used a novel distraction testing method that was introduced by Kujala and Mäkelä [17], to evaluate the visual distraction potential of the Nuviz HUD tasks and the smartphone tasks. One benefit of the method is that the in-vehicle glances made during the testing can be categorized to appropriate (green) or inappropriate (red) glances dependent on the visual demands of the road point where the glance begins.

The method follows the idea of Victor et al. [33] regarding the high change rate of the driving situation and Kircher and Ahlstrom's [11] idea about the timing of the off-road glance. Victor et al. [33] have noticed that many crashes occur because of a combination of a relatively short glance and high rate at which the dynamics of the driving situation changed during that glance – not because the off-road glance was too long as such. Also, Kircher and Ahlstrom [11] have suggested that all off-road glances are not equally distracting but timing of the off-road glance is critical.

In this study, glances towards the HUD view are interpreted as off-road (i.e., in-vehicle) glances since at least the driver's focal visual attention is then on the HUD. The appropriateness of an in-vehicle glance is determined here based on the visual demands of the route point and not solely based on the glance duration. Thus, the method may be more suitable to assess if the HUD affects drivers' situational awareness and glance timing than the NHTSA [25] recommended practice with static driving scenario and static glance acceptance criteria (see [29]).

HUD technology could cause a phenomenon called tunnel vision where the gaze is concentrated on too narrowly to the HUD and/or the road center [34], sacrificing observations for unexpected events in the road environment. Because of this, it is important to study also gaze concentration effects as a form of visual distraction particularly significant for HUDs. Victor, Harbluk and Engström [34] used a metric called percent road centre to measure these types of effects with in-vehicle tasks. This metric can tell if driver's field of view decreases compared to baseline driving. In this study, we analyze the gaze concentration effects by standard deviation of gaze in x-coordinates, which is another traditional measure used for analyzing differences in visual search patterns between novice and experienced drivers (e.g., [4]). The driving scenarios we utilized include turns in junctions, which stress the importance of horizontal observations for crossing traffic.

3 Method

For measuring the visual distraction potential of the in-vehicle tasks, we used a method by Kujala and Mäkelä [17]. The same method has been previously applied to study visual distraction potential of audio-visual route guidance [15] and different text entry methods [14]. This method utilizes visual occlusion technique, originally introduced by Senders, Kristofferson, Dietrich and Ward [28]. Visual occlusion refers to a condition where the driver's vision is occasionally occluded and the duration of the self-selected occlusion is measured. This is later referred as occlusion distance or OD. In this context, visual occlusion is used to measure the distance that is driven during the occluded period, not time. This enables the driver to freely control the driving speed during the measurement of the visual demands of driving.

The testing method is based on an experiment where 97 drivers' occlusion distances on simulated highway and suburban roads were measured [18]. These occlusion distances were mapped on the test routes and used during the distraction testing: the highway routes for participant sample validation and the suburban roads for the actual distraction testing. The participant sample validation by using their occlusion distances driven during an occlusion trial ensure that the driver sample includes both, "short-glancers" and "long-glancers". This validation is an important part of the testing method since previous studies have indicated that drivers have individual off-road glance duration tendencies [2, 16] and these individual differences in glance durations could affect the results of the distraction testing [3, 23].

During the distraction testing, the in-vehicle glance distances are measured. An in-vehicle glance refers to a glance that is directed to an in-vehicle device. Thus, an in-vehicle glance distance refers to a distance in meters that is driven during the in-vehicle glance. These in-vehicle glances can be categorized as green or red glances based on the original 97 drivers' occlusion data [18].

The categorization of the in-vehicle glances is based on the distance driven during a glance from a particular route point where the glance begins. A green glance refers to an in-vehicle glance length that is at or below the baseline data's median occlusion distance for the route point and therefore can be considered as an appropriate glance. The verification threshold for green glances has in previous studies [14, 15] been set to 68 % (min) of all the in-vehicle glances made during the task. To pass the verification criterion, the task should have 68 % or more of green glances. The verification criterion is based on the median percentage of the occlusion distances of the 97 drivers in the study of Kujala, Mäkelä, Kotilainen and Tokkonen [18].

A red glance refers to an in-vehicle glance length that exceeds the 85th percentile of the original 97-driver sample's occlusion distance on the route point. Red glances can thus be considered as inappropriately long in-vehicle glances in relation to the visual demand of the given driving situation. At these occasions, the in-vehicle task has caught the driver's attention for longer time than what the majority of the 97 drivers would have preferred to drive without vision on that route point. The verification threshold for red glances has been set to 6 % (max) [14, 15] of all the in-vehicle glances made during the task. If the task's red glances exceed 6 %, the task fails the verification criterion. The verification criterion is based on the 85th percentile of the occlusion distances of the 97-driver sample in the study of Kujala et al. [18].

3.1 Design of the study

The experimental design for the distraction testing was a within-subjects 2 x 3. The independent variables were the in-vehicle device: a helmet-mounted head-up display or a smartphone, and the in-vehicle task type: navigation, song search or phone call. For studying the gaze concentration effects of the Nuviz HUD (RQ3), the design was a within-subjects 2 x 1 (baseline driving versus route-following with the HUD). The design for NASA-TLX [8] was a within-subjects 4 x 1 the trial as an independent variable (baseline, Nuviz HUD tasks, smartphone tasks, occlusion).

3.2 Participants

The selection of the participants followed the NHTSA [25] recommendations regarding the driver sample for testing distraction of in-vehicle electronic devices, as precisely as possible. Convenience sampling was used to recruit participants via the University of Jyväskylä's mailing lists and connecting local motorcycle clubs. Altogether there were 24 participants (17 males and 7 females). The age of the participants varied from 19 to 72 years (M = 38.7; SD = 14.3). Six participants were 18 to 24 years old, seven were 25 to 39 years old, seven were 40 to 54 years old and four of the participants were older than 55 years. The driving experience varied from 1.5 years to 54 years (M = 20.6; SD = 15.0) and the driven kilometers per year from 6 000 to 40 000 (M = 16 542; SD = 10 283).

3.3 Apparatus

The experiments took place in a driving simulator laboratory of the University of Jyväskylä. A car simulator was used to conduct the experiment although this study is about testing a device that is designed for motorcycle drivers. The driving simulator (see Figure 1) is a medium-fidelity simulator with a motion platform (CKAS Mechatronics 2-DOF). The simulator has automatic transmission, force-feedback steering wheel and pedals (Logitech G27), and the seat is longitudinally adjustable. Eepsoft's (www.eepsoft.fi) professional driving simulator software was used for simulating the driving and saving driving log data at 10 Hz.

The simulator has three 40" LED screens (Samsung, 95.6 cm x 57.4 cm) with resolution of 1440 x 900 per screen. The screens display the driving scene as well as the rear-view mirror and side mirrors. We used a separate 7" tablet (Lenovo TB3-730X) above the steering wheel to display a speedometer to make the position of the meter resemble the meter position in a motorcycle. The tablet received the speed data in near real-time from the simulator software via a Wi-Fi network and the MockGeoFix Android application and displayed the speed to the participants with the Speedometer application available in PlayStore.

During the distraction testing, we used Samsung Galaxy A3 smartphone (4.5", Android 6.0.1), Nuviz head-up display that was mounted to a motorcycle helmet and a controller that was attached to the left side of a steering wheel (see Figure 1). The controller was positioned so that it could be used with the left-hand thumb without taking hands off the steering wheel. The controller is intended to be used in a similar manner when attached to a handlebar of a motorcycle. The smartphone was placed in a holder next to a steering wheel (see Figure 1). A laptop was used to mirror the Nuviz HUD image to ensure that the experimenter saw the same HUD view as the participants.

The Nuviz Android application was running in the same Samsung smartphone, with which the smartphone tasks were conducted. The application communicated with the Nuviz head-up display via a Bluetooth connection. The user interface of the Nuviz HUD can be seen in Figure 2. The functionalities shown in the upper and lower corners at the left side of the views could be selected by the right-hand buttons of the controller. The view (i.e., application or menu position) could be switched up or down by the central scroll button in the controller.

We recorded the eye movements with Ergoneers' Dikablis Essential 50 Hz head-mounted eye-tracking system and synchronized the driving simulator data with the eye-tracking data using LAN bridge.

For the occlusion trial, the steering wheel was outfitted with two levers behind the wheel that reveal the driving scene for 500 milliseconds per pull. If a lever is pulled repeatedly, the driving scene is visible all the time. The time of 500 milliseconds is based on the pioneering research on the occlusion method of Senders et al. [28].

During the trials, we used four predefined routes that simulate actual Finnish suburban roads in the Helsinki metropolitan area. All the routes used during the distraction testing were suburban roads without traffic. The same roads were used also in the study of Kujala and Mäkelä [17]. For the occlusion trial, we used a predefined highway route without any traffic. The driven route was same as for the baseline sample (N = 97) [18]. There were three speed limits: 60 kilometers, 80 kilometers and 120 kilometers per hour. The speed limit changed exactly at the same point for each participant.

3.4 Procedure

Demographic data was collected before the experiment via email. After arrival, the participants signed an informed consent form. At first, each participant adjusted the simulator's seat as close to the steering wheel as possible. This was done to make the HUD image appear above the road environment displayed on the middle screen. The distance between the seat and the steering wheel varied from 32 centimeters to 56 centimeters, mean being 45 centimeters.

Figure 1: Experimental setup and the position of the devices. The orange arrow points to the steering wheel-mounted controller and the green arrow points to the smartphone.

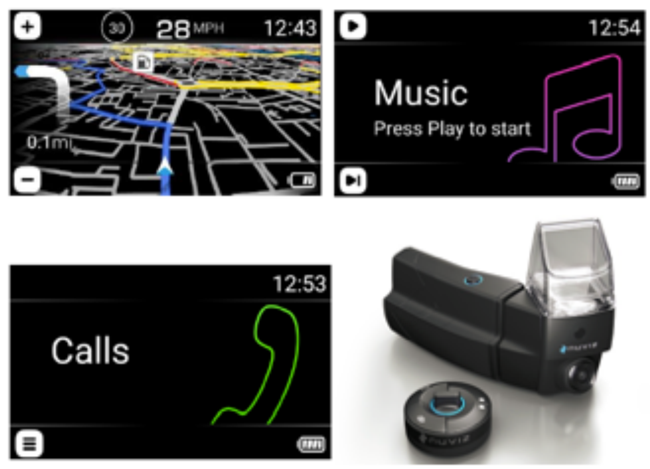

Figure 2: On top: Nuviz HUD user interfaces (navigation and music). On bottom: Nuviz HUD user interface (calls) and Nuviz HUD device with the controller. NB. During the distraction testing the controller did not have any labelling on it.

After adjusting the seat, the participants practiced driving with the simulator as long as they wanted. The practice driving scene was an artificial city environment with other road users. The average practice time was 4.19 minutes. When they started to be familiar with the simulator, they started to practice for the occlusion trial. The purpose of this practice was to get the participants familiar with driving occasionally without vision but still safely. The practice took place in the same artificial city environment with other traffic. The occlusion practice time was on average 6.67 minutes.

The first trial after the two practices was an occlusion trial. The occlusion trial is an important part of the participant sample validation in the testing method. In the occlusion trial all the screens were blank by default and by pulling either of the steering wheel's levers, participants were able to see the driving scene for 500 milliseconds per pull. During the trial, each participant was instructed to obey the traffic rules, to drive safely – but at the same time to try to drive without any visual information of the road (vision occluded) as long as they can.

An extra movie ticket was promised to those six participants who were able to drive the longest distances vision occluded but still accurately. This was done in order to get the participants to focus on the driving task but still trying to maximize the period when they drive without vision. After the trial, each participant filled out the NASA Task Load Index (TLX) questionnaire [8].

After the occlusion trial, the distraction testing started. At this point, the motorcycle helmet with the eye-tracker was put on and the eye-tracker was adjusted and calibrated.

After these preparations, the participants received general driving instructions: to prioritize driving task, to obey the traffic rules and to drive safely. They were also advised that the speed limit is 50 kilometers per hour but they were reminded that they may adjust the speed if needed.

All participants conducted three tasks with the smartphone and with the Nuviz HUD as well as a baseline drive where the task was to follow the verbal navigation instructions provided by the simulator software. The order of the tasks and the routes were counterbalanced in order to avoid learning effects. The baseline drive was always driven between the in-vehicle task trials. The visual demands of each used route were as similar as possible and there were no other road users or red traffic lights in the driving scenarios.

Each participant practiced to conduct the tasks with similar mock-up tasks before the actual task. The smartphone tasks were: 1) to follow the driving instructions to a destination, provided by Google Maps for a pre-defined route, 2) to find a target song from a list of unordered music and to start playing the song using Samsung's native music player (3 songs), and 3) to find a contact information from a list of contacts and to make a call using Samsung's native Contacts application and its call feature (3 calls). Neither of the latter tasks required typing, only scrolling the lists and selecting the correct song or contact information and to tap the call function.

The Nuviz HUD tasks were: 1) to follow the driving instructions to a destination provided by the Nuviz's user interface utilizing a predefined route on HERE maps installed on the Samsung phone, 2) to find a target song from a list of unordered music and to start playing the song using Nuviz's user interface that activated the Samsung's native music player (3 songs), and 3) to find a contact information from a list of contacts and to make a call using Nuviz's user interface that utilized the Samsung's contact information and call feature (3 calls). All the Nuviz tasks were conducted using the physical steering wheel controller that controlled the view in the head-up display. In other words, no typing was required, only scrolling the lists with the central scroll button and making selections by clicking an appropriate button in the corners of the controller.

After each task the participants filled out a reduced NASA-TLX questionnaire without weighting [8]. Finally, each participant was rewarded with a movie ticket.

3.5 Analyses

The main dependent variables for the distraction testing (RQ1) were the percentages of green and red in-vehicle glances and for the analysis of experienced task demands (RQ2) the total NASA-TLX score for each trial. We also report the total number of in-vehicle glances, as well as the total and mean glance duration, and the percentage of over-2-second in-vehicle glances, to provide comparable data with the NHTSA recommended verification criteria [25].

For the analysis of gaze concentration effects of the HUD (RQ3), we compared the standard deviation of the pupil's x-coordinate in eye camera pixels (i.e., horizontal gaze activity) as provided by the Ergoneers' D-Lab software (version 2.5), between the baseline driving and the navigation task with the Nuviz HUD.

The in-vehicle glance lengths were scored following the definition by SAE-J2396 [31]. However, the gaze transition time back to the driving scene was added to a glance duration to provide a 'full' off-road glance length. For the smartphone tasks, glances to the smartphone were counted as in-vehicle glances. For the Nuviz HUD tasks, glances to the HUD and the controller were counted as in-vehicle glances. During the testing, the in-vehicle glances were scored in real-time automatically with a script that recognized pupil's x and y coordinates provided by the Dikablis eye-tracking system. The coordinates were synchronized with the driving simulator's location data. All the automatically scored glances were manually reviewed from synchronized videos (25 fps) using Noldus XY software and all the inaccuracies were corrected frame-by-frame.

To ensure that our driver sample is compatible with the original driver sample [18], the range of the occlusion distances as well as median distances were measured. This was done is order to make sure that the use of the baseline occlusion data is appropriate and that there is no overrepresentation either in "short-glancers" or "long-glancers". Medians were chosen over means because of the non-Gaussian OD distributions. For controlling the effects of accelerations and decelerations in the beginning of the trial, in the junctions and in the end of the trial, only occlusion distances that were driven over 72 kilometers per hour (20 m/s) were included in the data.

Since the distributions of the green and red glances were also non-Gaussian, one-sample sign test was used to test the equality of the green and red glance percentages' medians to the verification thresholds (min 68 % green and max 6 % red). The differences between the Nuviz and smartphone tasks were analyzed with Wilcoxon signed-rank test. The non-parametric Wilcoxon signed-rank test was used also to compare the horizontal gaze activity between the baseline driving and the Nuviz HUD navigation task. For multiple pairwise comparisons, Bonferroni corrections were applied. Differences between the NASA-TLX scores were analyzed with Wilcoxon signed-rank test because most of the distributions were non-Gaussian. Where applicable, Cohen's *d* is reported as a measure of effect size.

4 Results

4.1 Occlusion distances

Due to technical problems in one trial, *N* of occlusion distances is 23. The occlusion distance varied from 9.0 to 37.4 meters (range 28.4 m), median being 20.7 meters. The equivalent range and median of the baseline data [18] are 3.2 - 41.9 meters (38.7 m) and 13.7 meters. Out of interest, there was a strong inverse correlation between occlusion distance and age: $r = -.50$.

4.2 Mean number of in-vehicle glances

Table 1 indicates that there were enough glances per task type for meaningful statistical testing. The mean numbers of glances for the song search and call tasks in Table 1 can be multiplied by three to get the total number of glances analyzed in the distraction testing.

Device	Navigation	Song search	Call
Nuviz HUD	74.50 (36.90)	19.24 (8.50)	8.78 (5.70)
Samsung smartphone	46.96 (21.47)	14.74 (11.65)	12.99 (7.97)

Table 1: Mean number of in-vehicle glances per task type. Standard deviation in parentheses. The song search and call tasks are averaged over three tasks.

4.3 Green in-vehicle glances

The verification threshold for green glances was set to 68 % (min) [14, 15]. According to one-sample sign test, the Nuviz HUD task passed the verification criterion for green glances (see Figure 3), the percentage being 68.64 ($p = .376$). The smartphone tasks did not pass the criterion since the median was 42.26 % ($p < .001$). Wilcoxon signed-rank test indicated that there is a significant difference in the green glance percentages between the Nuviz HUD and smartphone tasks: $Z = 3.49$, $p < .001$. The effect is large ($d = 0.90$).

Wilcoxon signed-rank test indicated that at task-level only statistically significant difference in green glance percentages between the devices, in favor of Nuviz HUD, was with the song search task: $Z = 3.71$, $p < .001$, $d = 1.39$ (large effect), with a Bonferroni-corrected alpha level of .017 (see Table 2). For the other two tasks, the pairwise differences were not statistically significant after Bonferroni-correction: navigation $Z = 2.29$, $p = .022$, $d = 0.75$; call $Z = 2.23$, $p = .026$, $d = 0.53$.

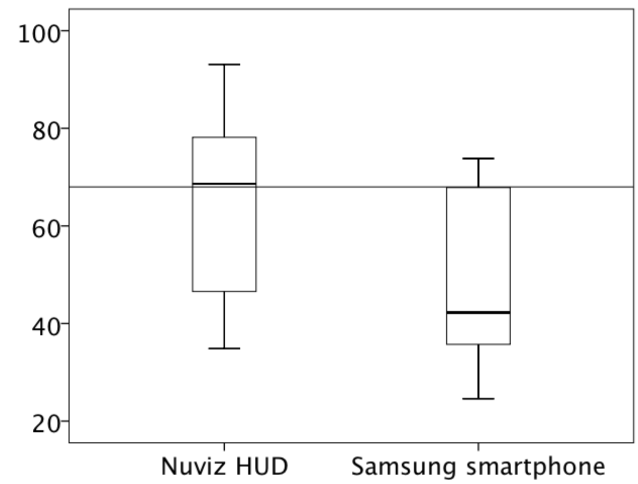

Figure 3: Percentage of green in-vehicle glances, verification threshold at 68 %.

Device	Navigation	Song search	Call
Nuviz HUD	56.6	73.6	68.3
Samsung smartphone	32.8	36.2	54.0

Table 2: Percentage of green in-vehicle glances (median) per input method and task type.

4.4 Red in-vehicle glances

The verification threshold of red glances was set to 6 % (max) [14, 15]. According to one-sample sign test, both – Nuviz HUD and smartphone – tasks passed the set verification criterion (see Figure 4). Nuviz's median red glance percentage was 3.41 ($p = .511$) and smartphone's 6.20 ($p = .162$). The smartphone's red glance percentage does not differ significantly from the threshold (6 %) and therefore passed the test.

According to Wilcoxon signed-rank test, there is also a significant difference in the red glance percentages between the Nuviz HUD and smartphone tasks: $Z = 2.74$, $p = .006$. The effect is of medium size ($d = 0.62$).

Wilcoxon signed-rank test indicated that there was a significant difference between the devices (see Table 3), in favor of Nuviz HUD, in the song search task ($Z = 2.66$, $p = .008$, $d = 0.82$ [large effect]) and in the navigation task ($Z = 2.40$, $p = .016$, $d = 0.57$ [medium effect]), with the Bonferroni-corrected alpha level of .017. No difference was found in the call task ($Z = .501$, $p = .616$).

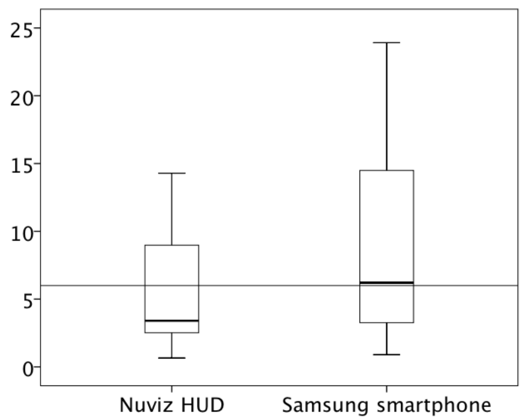

Figure 4: Percentage of red in-vehicle glances, verification threshold at 6 %.

Device	Navigation	Song search	Call
Nuviz HUD	3.8	2.0	4.1
Samsung smartphone	9.1	8.0	3.2

Table 3: Percentage of red in-vehicle glances (median) per input method and task type.

4.5 Durations of in-vehicle glances (NHTSA, 2013)

For enabling comparison between studies, also the NHTSA [25] recommended metrics are reported in Tables 4-6.

	Navigation		Song search		Call	
Device	M (SD)	85th %ile	M (SD)	85th %ile	M (SD)	85th %ile
Nuviz HUD	56.29 (17.69)	74.88	14.30 (3.21)	17.06	6.88 (2.32)	10.23
Samsung smartphone	50.56 (20.67)	77.65	16.22 (9.21)	26.12	11.41 (4.02)	15.87

Table 4: Total duration of in-vehicle glances (s). Standard deviation in parentheses. The song search and call tasks are averaged over three tasks.

	Navigation		Song search		Call	
Device	Md	85th %ile	Md	85t %ile	Md	85th %ile
Nuviz HUD	0.0	3.5	0.0	2.1	0.0	5.8
Samsung smartphone	0.0	5.9	8.6	25.3	4.7	12.8

Table 5: Percentage of over-2-second in-vehicle glances (median). Percentages calculated for three tasks per task type for the song search and call tasks.

	Navigation		Song search		Call	
Device	M (SD)	85th %ile	M (SD)	85th %ile	M (SD)	85th %ile
Nuviz HUD	0.86 (0.27)	1.13	0.81 (0.23)	1.09	0.87 (0.25)	1.14
Samsung smartphone	1.13 (0.22)	1.33	1.23 (0.35)	1.60	1.00 (0.35)	1.38

Table 6: Mean in-vehicle glance durations (s). Means calculated for three tasks per task type for the song search and call tasks.

4.6 Experienced task workload - NASA-TLX

According to Wilcoxon signed-rank test, all the differences between the trials were significant with $\alpha = .008$, except the difference between occlusion trial and smartphone tasks ($p = .158$, see Figure 5): baseline vs. Nuviz HUD tasks, $Z = 3.42$, $p = .001$, $d = 0.73$; baseline vs. occlusion, $Z = 4.17$, $p < .001$, $d = 2.20$; baseline vs. smartphone tasks, $Z = 4.26$, $p < .001$, $d = 1.78$; Nuviz HUD tasks vs. occlusion, $Z = 3.93$, $p < .001$, $d = 1.36$; and Nuviz HUD tasks vs. smartphone tasks, $Z = 3.33$, $p = .001$, $d = 1.02$. All the effect sizes are large, except between baseline driving and Nuviz HUD tasks the effect size is medium.

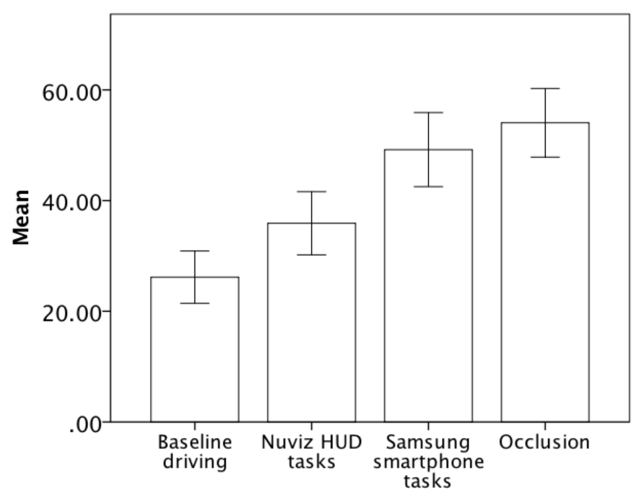

Figure 5: Experienced task workload measured with NASA-TLX. Maximum is 100. Error bars: 95 % CI.

4.7 Horizontal gaze activity

The horizontal gaze activity was measured with standard deviation of pupil's x-coordinate in eye camera pixels (see Figure 6). According to Wilcoxon signed-rank test, there was no significant difference in horizontal gaze activity between the baseline driving and the Nuviz HUD navigation task ($p = .349$).

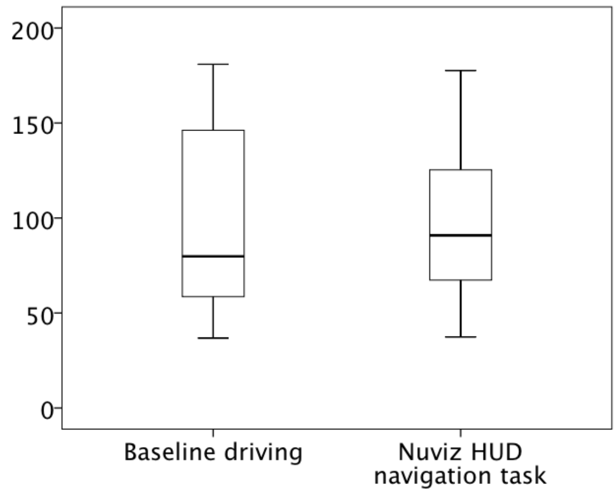

Figure 6: Standard deviation of pupil's x-coordinate in eye camera pixels.

5 Discussion

This was the first study comparing visual distraction effects between tasks conducted with a helmet-mounted HUD for motorcyclists and with a smartphone. The distraction potential of these two devices were assessed with red and green in-vehicle glances as defined by Kujala and Mäkelä [17]. The verification threshold for green glances was set to 68 % (min) and for red glances maximum to 6 % (max) [14, 15]. Overall, the Nuviz HUD tasks passed the set verification criterion for green glances, the percentage of green glances being 68.6. The smartphone tasks did not pass this criterion with the percentage of 42.3. When compared per task type, the song search task conducted with Nuviz HUD had significantly higher green glance percentages than the similar task with a smartphone (here, higher is better).

Both – Nuviz HUD and smartphone tasks – passed the set verification criterion for the inappropriately long red glances. The overall percentage for the Nuviz HUD tasks was 3.4 and for the smartphone tasks 6.2. The smartphone tasks passed the criterion as well because the difference between the percentage and the verification threshold is not significant. However, the difference in red glance percentages between the Nuviz HUD and smartphone tasks is significant. When compared per task type, the song search and navigation tasks had significantly lower red glance percentages when conducted with the Nuviz HUD.

Based on these findings, the studied Nuviz HUD tasks seem to have lower distraction potential than the tasks with the same goals conducted with an Android smartphone (RQ1). Compared to the smartphone tasks, the Nuviz HUD tasks increased the percentage of green in-vehicle glances by 62 % and decreased the percentage of red in-vehicle glances by 45 %. Also, there was no overrepresentation of "short-glancers" or "long-glancers" in the participants' occlusion distance distribution. Based on that, the sample can be considered comparable with the baseline 97-driver sample [18] and the green and red glance metrics as reliable.

The experienced task workload was reported highest during the occlusion trial and second highest during the smartphone tasks. No significant difference between these two trials were found. The baseline drive was experienced as the least demanding and the Nuviz HUD tasks were experienced as second least demanding. There was a significant difference between baseline driving and Nuviz HUD tasks as well as between Nuviz HUD tasks and smartphone tasks. The Nuviz HUD tasks were experienced less demanding than the smartphone tasks (RQ2).

We found no difference in horizontal gaze activity between baseline driving and the navigation task conducted with the Nuviz HUD. It can be interpreted that using HUD for navigation while driving did not cause gaze concentration in this study (RQ3). The possible gaze concentration effects of HUDs should be further studied also with other types of in-vehicle tasks than navigation.

Due to confounding factors, we cannot pinpoint the exact design factors explaining the advantage of the tasks with the Nuviz HUD over the similar tasks with a smartphone. However, based on earlier research, we assume that the HUD enabled use of peripheral vision to maintain better situational awareness of the demands of the driving environment during the HUD tasks compared to the head-down phone tasks. In addition, the easy-to-learn haptic design of the HUD controller could be used without vision. We noticed that glances directed to the steering wheel-mounted HUD controller were very rare, from a few to none. This is a positive sign towards the haptic design of the controller.

Previous studies (e.g., [35, 36]) have found similar results concerning HUDs operated with physical controllers and touchscreen HDDs. However, the tested Nuviz HUD differs from those since it is designed to be used while driving a motorcycle and the HUD is helmet-mounted, ensuring that the HUD view is visible for the driver in all head positions. This is not the case with windshield HUDs [35, 36]. Windshield HUDs [35, 36] cannot be used in a motorcycle context since motorcycles are missing car-like windshields and that is why a different HUD solution is needed.

As an example of UI design differences at the software level, both music search tasks required scrolling a list of music to find the target song(s). With the Nuviz HUD the participant could scroll the list song-by-song by a single press of a physical button in the steering wheel controller. With the smartphone, the participant had to point the touchscreen and scroll the music player menu by means of kinetic scrolling, which has been found to be one of the most visually distracting activities with touchscreen in-vehicle devices [19, 20]. In addition, the participant did not have to look down, far away from the driving environment, to see the selected song in the HUD.

The navigation, song search, and call tasks had equivalent or even higher mean number and total duration of in-vehicle glances conducted with the Nuviz HUD than with the smartphone (Table 1). Despite that, the Nuviz HUD tasks had higher green glance percentages and lower red glance percentages than the tasks conducted with the smartphone. This finding suggests, that these metrics of visual demand may suit poorly for measuring visual distraction. This is the case, in particular, if visual distraction is operationalized as a calibration failure between the situational visual demands of driving and the off-road glance length. Similarly, Smith, Gabbard and Conley [29] found out that HUD affected more negatively the NHTSA glance metrics [25] than HDD. Depending on the user interface, the task length may not be as critical factor for appropriate timing of the in-vehicle glances and distraction than other task features. The finding stresses the importance of in-vehicle user interface and task design to mitigate visual distraction and the importance of using proper metrics suitable for a user interface design in distraction testing. The green and red glance percentages observed in this study are well in line with previous findings [14, 15]. This gives credibility to the used distraction testing method as similar task designs seem to produce similar results.

Nonetheless, there are some limitations concerning this study. The car simulator used in this study cannot simulate driving with a motorcycle. In addition, in this study the Nuviz controller was attached to the steering wheel (see Figure 1). When driving a motorcycle, the controller would be attached to a handlebar. However, the study was designed to enable comparative analysis of the visual distraction effects of device used while driving (a simulated car) in a controlled environment. Motorcycling can be argued to be more demanding than driving a car [24], and thus, the absolute distraction effects may be even larger while riding a motorcycle than what measured here. Naturally this applies to both smartphone tasks and HUD tasks. The measured visual distraction effects of the in-vehicle tasks cannot be generalized to provide estimates of the absolute distraction effects while driving a motorcycle, but we argue that the observed relative effects between the devices and tasks are reliable. In fact, the generalization of any driving performance or glance data measured in a simulator to real conditions has to be done with caution.

Road surface roughness was absent because the simulator's motion platform has only two degrees of freedom. This factor could favor the Nuviz HUD with the thumb-controller even more in real traffic conditions. All the in-vehicle tasks in this study, also with the smartphone, were relatively easy due to low number of task steps. This was due to the fairly limited Nuviz HUD functionalities at the time of testing. With more complex in-vehicle tasks the distraction effects of both smartphone and HUD tasks could be worse. One should also keep in mind the usability-distraction paradox: the overall distraction effects in a driver population may be increased by safer and easier-to-use in-vehicle user interfaces, if these increase the frequency of use of these devices on the roads.

6 Conclusion

This was the first research comparing visual distraction effects of a HUD designed for motorcyclists to distraction effects of smartphone usage. The distraction effects were evaluated with a novel method that classifies in-vehicle glances to appropriate and inappropriate glances dependent on the situational driving demands. Compared to the smartphone tasks, the Nuviz HUD tasks increased the percentage of acceptable in-vehicle glances by 62 % and decreased the percentage of inappropriately long in-vehicle glances by 45 %. Based on the results, the tested HUD tasks seem to be safer for motorcyclists than similar tasks with a smartphone while driving.

The tasks conducted with the HUD where also reported less demanding than the tasks conducted with the smartphone. The use of the HUD for navigation guidance did not cause gaze concentration effects compared to baseline driving. However, these effects are something to be studied more carefully in the future with other types of in-vehicle tasks.

The study had some limitations since a car simulator was used instead of a motorcycle simulator. On the other hand, driving a motorcycle is more complex task than driving a car and that is why the distraction effects can be even larger than what reported while riding a motorcycle.

ACKNOWLEDGMENTS

We thank Sara Norja for conducting the experiments. The study was partly funded by Nuviz Inc. The funder had no role in data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

[1] Stewart A. Birrell and Mark S. Young. 2011. The Impact of Smart Driving Aids on Driving Performance and Driver Distraction. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(6), 484–493. DOI: https://doi.org/10.1016/j.trf.2011.08.004

[2] Robert Broström, Peter Bengtsson and Mikael Ljung Aust. 2016. Individual Glance Strategies and Their Effect on the NHTSA Visual Manual Distraction Test. *Transportation Research Part F: Traffic Psychology and Behaviour*, 36, 83-91. DOI: http://dx.doi.org/10.1016/j.trf.2015.10.017

[3] Robert Broström, Mikael Ljung Aust, Linnea Wahlberg and Laban Källgren. 2013. What Drives Off-Road Glance Durations During Multitasking – Capacity, Practice or Strategy. In *Proceedings of the 3rd International Conference on Driver Distraction and Inattention*.

[4] David E. Crundall and Geoffrey Underwood. 1998. Effects of Experience and Processing Demands on Visual Information Acquisition in Drivers. *Ergonomics*, 41, 448-458. DOI: https://doi.org/10.1080/001401398186937

[5] Department of Environment, Transport and the Regions (DETR) Report. 2000. Tomorrow's Roads—Safer for Everyone: The Government's Road Safety Strategy and Casualty Reduction Targets for 2010, HMSO.

[6] Gregory M. Fitch, Susan A. Soccolich, Feng Guo, Julie McClafferty, Youjia Fang, Rebecca L. Olson, Miguel A. Perez, Richard J. Hanowski, Jonathan M. Hankey and Thomas A. Dingus. 2013. The Impact of Hand-Held and Hands-Free Cell Phone Use on Driving Performance and Safety-Critical Event Risk. *Report No. COT HS 811 757. National Highway Traffic Safety Administration, U.S. Department of Transportation.*

[7] James P. Foley, Richard Young, Londa Angell and Joshua E. Domeyer. 2013. Towards Operationalizing Driver Distraction. In *Proceedings of the Seventh International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design.* DOI: 10.17077/drivingassessment.1467

[8] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in psychology*, 52, 139-183. DOI: https://doi.org/10.1016/S0166-4115(08)62386-9

[9] Jibo He, William Choi, Jason S. McCarley, Barbara S. Chaparro and Chun Wang. 2015. Texting while Driving Using Google Glass™: Promising but not Distraction-Free. *Accident Analysis & Prevention*, 81, 218-229. DOI: https://doi.org/10.1016/j.aap.2015.03.033

[10] Jibo He, Jason S. McCarley, Kirsten Crager, Murtuza Jadliwala, Lesheng Hua and Sheng Huang. 2018. Does Wearable Device Bring Distraction Closer to Drivers? Comparing Smartphones and Google Glass. *Applied Ergonomics*, 70, 156-166. DOI: https://doi.org/10.1016/j.apergo.2018.02.022

[11] Katja Kircher and Christer Ahlstrom. 2017. Minimum Required Attention: A Human-Centered Approach to Driver Inattention. *Human Factors*, 59(3), 471-484.

[12] Sheila G. Klauer, Thomas A. Dingus, Vicki L. Neale, Jeremy D. Sudweeks and D. J. Ramsey. 2006. The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data. *National Highway Traffic Safety Administration, U.S. Department of Transportation.*

[13] Sheila G. Klauer, Feng Guo, Bruce G. Simons-Morton, Marie Claude Ouimet, Suzanne E. Lee and Thomas A. Dingus. 2014. Distracted Driving and Risk of Road Crashes among Novice and Experienced Drivers. *New England Journal of Medicine*, 370(1), 54-59.

[14] Tuomo Kujala and Hilkka Grahn. 2017. Visual Distraction Effects of In-Car Text Entry Methods: Comparing Keyboard, Handwriting and Voice Recognition. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications ACM,* 1-10. DOI: https://doi.org/10.1145/3122986.3122987

[15] Tuomo Kujala, Hilkka Grahn, Jakke Mäkelä and Annegret Lasch. 2016. On the Visual Distraction Effects of Audio-Visual Route Guidance. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, *ACM*, 169-176. DOI: https://doi.org/10.1145/3003715.3005421

[16] Tuomo Kujala, Annegret Lasch and Jakke Mäkelä. 2014. Critical Analysis on the NHTSA Acceptance Criteria for In-Vehicle Electronic Devices. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 1-8. DO: https://doi.org/10.1145/2667317.2667322

[17] Tuomo Kujala and Jakke Mäkelä. 2015. Development of a Testing Environment and a Verification Procedure for In-Vehicle Tasks with Dynamic Driving Scenarios. In *Proceedings of the 4th International Conference on Driver Distraction and Inattention*.

[18] Tuomo Kujala, Jakke Mäkelä, Ilkka Kotilainen and Timo Tokkonen, 2016. The Attentional Demand of Automobile Driving Revisited: Occlusion Distance as a Function of Task-Relevant Event Density in Realistic Driving Scenarios. *Human Factors*, 58(1), 163-180. DOI: https://doi.org/10.1177/0018720815595901

[19] Tuomo Kujala, Johanna Silvennoinen and Annegret Lasch. 2013. Visual-Manual In-Car Tasks Decomposed: Text Entry and Kinetic Scrolling as The Main Sources of Visual Distraction. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 82-89. DOI: https://doi.org/10.1145/2516540.2516562

[20] Annegret Lasch and Tuomo Kujala. 2012. Designing Browsing for In-Car Music Player: Effects of Touch Screen Scrolling Techniques, Items per Page and Screen Orientation on Driver Distraction. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 41-48.

[21] Felix Lauber, Claudius Böttcher and Andreas Butz. 2014. You've Got the Look: Visualizing Infotainment Shortcuts in Head-Mounted Displays. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 1-8. DOI: https://doi.org/10.1145/2667317.2667408

[22] Ja Young Lee, Madeleine C. Gibson and John D. Lee. 2016. Error Recovery in Multitasking While Driving. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5104-5113. DOI: https://doi.org/10.1145/2858036.2858238

[23] Ja Young Lee and John D. Lee. 2017. Multi-level Analysis of Distracted Drivers' Glances: Enhancing the Robustness of the NHTSA Acceptance Criteria. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 61(1), 1919-1923.

[24] Fred L. Mannering and Lawrence L. Grodsky. 1995. Statistical Analysis of Motorcyclists' Perceived Accident Risk. *Accident Analysis & Prevention*, 27(1), 21-31.

[25] National Highway Traffic Safety Administration. 2013. *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices*. (NHTSA-2010-0053.) Washington DC, NHTSA.

[26] Rebeca L. Olson, Richard J. Hanowski, Jeffrey S. Hickman and Joseph Bocanegra. 2009. Driver Distraction in Commercial Vehicle Operations. *Report No. FMCSARRR-09-042. Federal Motor Carrier Safety Administration, U.S. Department of Transportation.*

[27] Sysavanh Phommachanh, Masao Ichikawa, Shinji Nakahara, Mayfong Mayxay and Akio Kimura. 2017. Student Motorcyclists' Mobile Phone Use while Driving in Vientiane, Laos. *International Journal of Injury Control and Safety Promotion*, 24(2), 245-250. DOI: https://doi.org/10.1080/17457300.2016.1166141

[28] John W. Senders, A. B. Kristofferson, W. H. Levison, C. W. Dietrich and J. L. Ward. 1967. The Attentional Demand of Automobile Driving. *Highway Research Record*, 195.

[29] Missie Smith, Joseph L. Gabbard and Christian Conley. 2016. Head-Up vs. Head-Down Displays: Examining Traditional Methods of Display Assessment while Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, 185-192. DOI: https://doi.org/10.1145/3003715.3005419

[30] Missie Smith, Jillian Streeter, Gary Burnett and Joseph L. Gabbard. 2015. Visual Search Tasks: The Effects of Head-Up Displays on Driving and Task Performance. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 80-87. DOI: https://doi.org/10.1145/2799250.2799291

[31] Society of Automotive Engineers. 2000. *SAE-J2396 Definitions and Experimental Measures Related to the Specification of Driver Visual Behavior Using Video Based Techniques*. Warrendale, PA: SAE.

[32] Long T. Truong, Chris De Gruyter and Hang T. T. Nguyen. 2017. Calling, Texting, and Searching for Information while Riding a Motorcycle: A Study of University Students in Vietnam. *Traffic Injury Prevention*, 18(6), 593-598. DOI: https://doi.org/10.1080/15389588.2017.1283490

[33] Trent Victor, Marco Dozza, Jonas Bärgman, Christian-Nils Boda, Johan Engström, Carol Flannagan, John D. Lee, and Gustav Markkula. *Analysis of Naturalistic Driving Study Data: Safer Glances, Driver Inattention, and Crash Risk.* Washington, DC: Transportation Research Board.

[34] Trent Victor, Joanne L. Harbluk and Johan A. Engström. 2005. Sensitivity of Eye-Movement Measures to In-Vehicle Task Difficulty. *Transportation Research Part F: Traffic Psychology and Behaviour*, 8(2), 167-190. DOI: https://doi.org/10.1016/j.trf.2005.04.014

[35] Gabriella Villalobos-Zúñiga, Tuomo Kujala and Antti Oulasvirta. 2016. T9+ HUD: Physical Keypad and HUD Can Improve Driving Performance while Typing and Driving. In *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications,* 177-184. DOI: https://doi.org/10.1145/3003715.3005453

[36] Garrett Weinberg, Bret Harsham and Zeljko Medenica. 2011. Evaluating the Usability of a Head-Up Display for Selection from Choice Lists in Cars. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications, ACM*, 39-46.