Recognition of Text and Shapes on a Large-Sized Head-Up Display

Renate Haeuslschmid
LMU Munich
Munich, Germany
renate.haeuslschmid@ifi.lmu.de

Susanne Forster
IAV GmbH
Munich, Germany
susanne.forster@iav.de

Katharina Vierheilig
IAV GmbH
Munich, Germany
katharina.vierheilig@iav.de

Daniel Buschek
LMU Munich
Munich, Germany
daniel.buschek@ifi.lmu.de

Andreas Butz
LMU Munich
Munich, Germany
andreas.butz@ifi.lmu.de

ABSTRACT

The ever-increasing amount of information in cars demands novel and safer displays, such as head-up or windshield displays. We present two studies that investigate the recognition of stimuli presented on a windshield display. Using a divided attention task and a driving simulator, we first compared four types of stimuli which are common in traffic signs: text, circles, triangles, and squares. We measured the response times at 17 positions within an extended field of regard of $35° \times 15°$. The follow-up study validated our results by replicating the first study with two changes: We investigated the influence of peripheral workload with a more diverse simulated environment and tested for training effects by converting the setup to a left-hand drive car. We contribute response times and sizing recommendations for a field of regard of $35° \times 15°$. These recommendations will help designers of large head-up displays to create interfaces which are well-legible and avoid both cluttering the driver's view and occluding the road scene.

Author Keywords

Head-up display; windshield display; in-vehicle interfaces; sizing recommendations; interface guidelines

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation (e.g. HCI): User Interfaces

INTRODUCTION

Head-up displays (HUDs) are about to become a standard feature in upper-class cars. These small displays or projection systems are reflected in the windshield to appear as a transparent display floating above the cars' hood in approximately 2 m distance to the driver. The size of these HUDs is still fairly small and hence the presentation of information is limited to a small area below the driver's line of sight. Extensive research and development is undertaken to increase the display size and cover larger parts of the driver's field of regard, potentially up to the entire windshield [7]. With an increased size, such displays offer many opportunities to display additional information in a safer or more intuitive way, for example by using augmented reality for hazard warnings [8]. These hazards might appear not only in front of the driver (e.g., emergency braking of the leading car) but also in the periphery (e.g., a pedestrian crossing the road), which is generally associated with less perceptual and cognitive resources [9] – leading to slower reactions. Since reacting in time can be crucial in a safety-critical task such as driving, it is important to display information in such a way that the driver's response times are minimized. For a fast response, both size and saliency of the visual information plays an important role [1, 13, 16, 23].

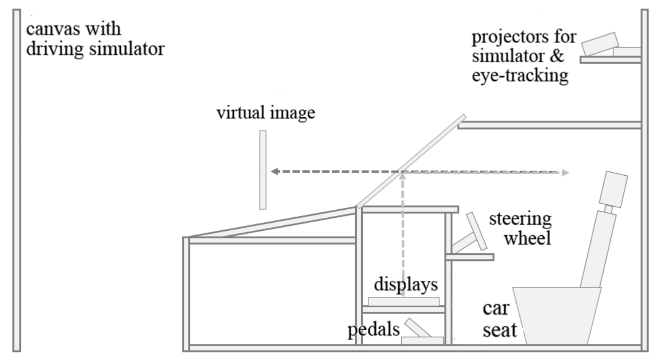

Figure 1. The test setup used for the two studies was nearly identical. The wooden construction matches the dimensions of a small car and creates a comparable view. The three 32" displays are lined up below the windshield so that their image is reflected on it and appears floating above the car's hood.

In this paper, we present two studies that investigate the driver's response times and suitable minimal sizes of windshield display (WSD) information (Figure 1). We applied the standard lane-change task (LCT) as a primary task and a detection-response task (DRT) as a secondary task. In this

DIS 2017, June 10–14, 2017, Edinburgh, United Kingdom.

http://dx.doi.org/10.1145/3064663.3064736

divided attention situation, we tested 17 different positions in a visual field of $35° \times 15°$, and four different types of stimulus, namely text and three shapes. In addition, we compared two different driving scenes as well as driving in a left-hand and a right-hand drive car. We translate these results into sizing recommendations which will help designers to choose sizes and locations for displayed information, such that it is both legible and does not cover the driving scene.

BACKGROUND & RELATED WORK

The Driver's Perceptual Capabilities

About 90% of the driving task requires the visual channel [2]. Hence, the driver's visual perception plays a crucial role for safety. However, visual perception is not a constant on which designers of in-car applications can rely. It depends on several interfering and uncontrollable factors such as the (primary) task demand, the (secondary) task location, the visual resolution, the driving speed, and the current state of the driver, such as drowsiness, anxiety or mental load [3, 4, 9, 20]. Moreover, visual perception also varies considerably across age groups and individuals. In particular, these factors influence the driver's useful or functional field of view (FoV), which comprises the area that is easy to track for the driver and of which he or she is constantly aware [8]. A reduction of this FoV corresponds to the well-known tunnel vision effect.

We further distinguish the central FoV ($<2°$), in which humans see sharp and show low response times, the foveal FoV($<10°$), which provides good contrast, contours, and color perception, and the peripheral FoV ($>10°$), which is limited to motion, light alteration, and orientation of objects [5, 7, 17]. In addition, there is evidence that humans can identify stimuli appearing even far off the line of sight and gather information about spatial characteristics (e.g., shape and position) [9], provided its design is adjusted: a larger size, a higher contrast, the choice of specific colors as well as motion or animation can promote recognition of peripheral information but can also increase distraction [1, 13, 16, 23].

Overall, perception deteriorates from the center (0°) towards the boundaries of the overall FoV. This suggests to display information in a rather central position to enable low access and response times. However, occlusion of important parts of the driving scene leads to considerably slower response times in the primary task. Hence, placing information in the foveal area is avoided and limited to the lower (HUD) area within the central FoV. However, information exceeding the foveal area – such as world-related information – will inevitably make use of the periphery. For peripheral stimuli research shows divergent results from almost no increase in response time to failure to detect the stimuli at all, as explained below.

Location-dependent Stimulus Recognition

So far, extensive research has been performed to find the most suitable location for head-up displays; Gish and Staplin [6] provide a good summary on this topic. These studies usually compare several collateral positions at the level of the horizon and below. Simple detection-response tasks – challenging the participant to respond by button presses to stimuli appearing at these locations – have been applied to measure secondary task response times while performing a primary driving task. These studies agree that the best location for a head-up display is at 5° below the driver's line of sight as response times and road occlusion are low for this location.

However, since the aim of these studies was to identify only one location for a fairly small display, they investigated only a very limited windshield area: Tsimhoni et al. [22] tested 15 locations within an area of 10° horizontally and 5° vertically. The best secondary task performance was reported for the central position (0°) and 5° to each side. In a similar study, Tsimhoni [21] tested locations within a field of 21.6° horizontally and 7.4° vertically and found that response times on secondary task events increased with higher eccentricity (distance from the center) while detection time remained quite constant (non-significant). According to Lino et al. [15], the secondary task performance is impeded when the task is placed 20° off a highly demanding primary task. Further, tasks placed more than 30° away from the primary tasks are found to degrade primary task performance [11]. This is in line with the results found by Lamble et al. [14]. Since no larger angles have been tested in these studies, it remains a subject of speculation how primary and secondary task performance will change at higher eccentricities and how they relate to task difficulty.

Since research & development now aims to enlarge the HUD size, it is important to investigate the response times outside of the area tested so far. The response times across the windshield will serve as an indicator of how to use the new display area: Critical information should be superimposed at locations which promote fast responses to stimuli while other content (e.g., entertainment functions) will not be harmed by increased response times. Furthermore, comparably little research has been devoted to the presentation of content within the driver's peripheral FoV. Information that is bound to the world, such as hazard warnings, may appear far off the driver's line of sight but is still critical for driving safety. Hence, response times within the peripheral FoV are also of high relevance for future WSD interface concepts.

STUDY 1: LOCATION-DEPENDENT RESPONSE TIMES

In the first user study we applied a lane change test and a detection-response task. Participants had to recognize stimuli which appeared at 17 positions across the windshield display. Each stimulus appeared slowly by growing in size which allows to derive a well-legible size from the response times.

Hypotheses:

H1.1: Reaction time increases with the visual angle of a stimulus.

H1.2: There are no differences in reaction time for the shapes *triangle*, *square*, and *circle*.

H1.3: Reaction times are higher for textual stimuli compared to shapes.

Independent Variables:

Stimulus Type: Either *text*, *triangle*, *square*, or *circle*.

Position: The visual angle at which a stimulus appears (in °).

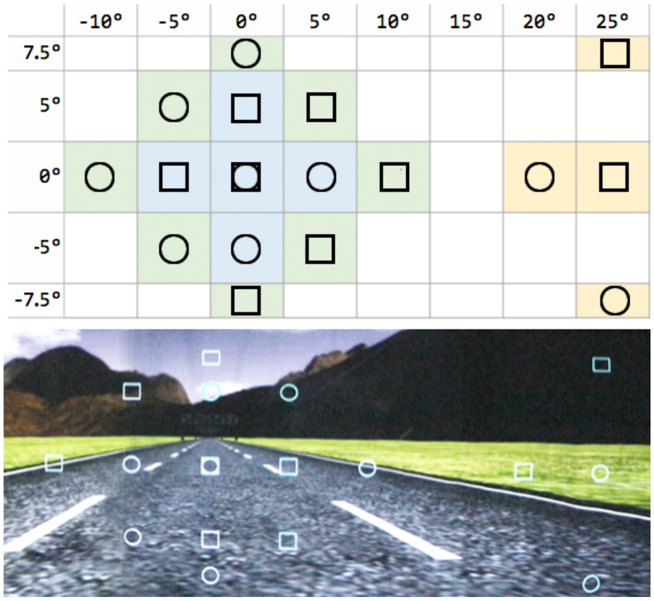

Figure 2. These 17 positions were tested in both studies. We chose this pattern as the best compromise between a low number of positions and a wide range of angles. We could not test a position at 15° due to limitations of the WSD setup. The shapes at the single positions represent the type of stimulus assigned to this position in the second study. In the first study, each type of stimulus was displayed at each position. The background color of the cells represents the field of view for the analysis of the results(<5° = blue; <15° = green; >15° = yellow). The lower image shows how the positions relate to the driving simulator scene.

Dependent Variables (Measurements):

Driving Performance: We evaluate the driving performance by measuring the mean deviation, the divergence between optimal lane position and actually driven lane position (in meters), and the reaction time, namely the time between the displayed lane change trigger and the initiation of a steering reaction by the driver (in milliseconds). We further report the successfully performed lane changes as success rate. *Detection-Response*

Performance: We evaluate the detection-response performance by calculating the response time; the time between the onset of the stimuli and the driver's button press. Due to the pre-defined scaling of the stimuli, the response time can be translated into a well-legible stimulus size (in degrees). By the word size we refers to the height of the text and to the height as well as width of the shapes.

Participants

We recruited 24 participants (7 female) with an average age of 30 years (SD=8.7). All participants owned a valid drivers' license and had normal or corrected-to-normal vision. Their average height was 180 cm (SD=10.7 cm).

Study Design

The lane change test is a standardized driving task invented for an easy and effective measurement of the distraction and influence of secondary or tertiary tasks on driving performance; developed by BMW and Daimler Chrysler [10]. The LCT asks the driver to change lanes on an empty three lane highway. The lane changes are randomized but positioned in a standardized frequency along the test track: each test track is 3.000 m long and demands a lane change every 150 m; corresponding to 18 lane changes. In our test, those lane changes are triggered by green, filled circles placed on overhead gantries, above the target lane. Each test track lasted for three minutes due to the fixed speed of 60 km/h.

The described driving task is constant throughout all baseline and intervention tracks. The test begins and ends with a baseline track and alternates with intervention tracks inbetween to compensate for learning or fatigue effects. This amounts to 5 baseline and 4 intervention tracks. In intervention tracks, the driver is additionally challenged by performing a secondary or tertiary task – in our case a detection-response task: The detection-response task requires participants to detect and respond to stimuli appearing in their visual field, for example by button presses [18].

We decided for a between-subjects design with two groups in order to increase the number of positions with respect to the LCT track duration and the timing of stimuli. Each group experienced 18 stimuli per test track – one target and one confounding stimulus at nine different positions (center is included in both groups). The participants had to react only to the target stimuli by pressing one predefined button and ignore the confounding stimuli. We integrated confounding stimuli to force the driver to consciously recognize and decide on the stimulus type but did not request an explicit action for the confounding stimuli in order to reduce the required buttons and the mental (re-)mapping of the stimulus types to buttons.

We decided to test 17 windshield display positions as this number allowed an acceptable compromise between an appropriate coverage of the windshield area and the number of conditions. The 17 positions are spread within a field of view of -10° to 25° horizontally and -7.5° to 7.5° vertically. The pattern of the 17 positions is depicted in Figure 2. It is more dense in the central area around the drivers' line of sight since information is generally recommended to be placed close to the drivers' line of sight to enable a fast response, as discussed by Haeuslschmid et al. [7]. The three display setup did not allow to place any stimulus at 15° due to the display frames.

As for the stimuli, it was important to use meaningless stimuli (to avoid a bias due to varying relevance) which do not require a high resolution or large size to be recognizable. We chose three shapes which can be found in common traffic signs: triangle, rectangle, and circle. In addition, we decided to test text: Five short and comparable male and female names (e.g., Eva and Tom) were selected, as proposed by Tsimhoni et al. [22]. To ensure that the text is well-legible, we used a sans serif font with large internal spaces and without special text formatting. To obtain shapes which are comparable to the relatively gracile text, we decided to display only the outline of the shapes. All stimuli were presented in the neutral and well distinguishable color cyan.

The stimuli are displayed at a non-recognizable size (10% of the maximum size) and slowly but continually scaled up to a well-legible size: the maximum height for text is 1.7° and 1.2° for forms. We evaluated these values in a pre-study to

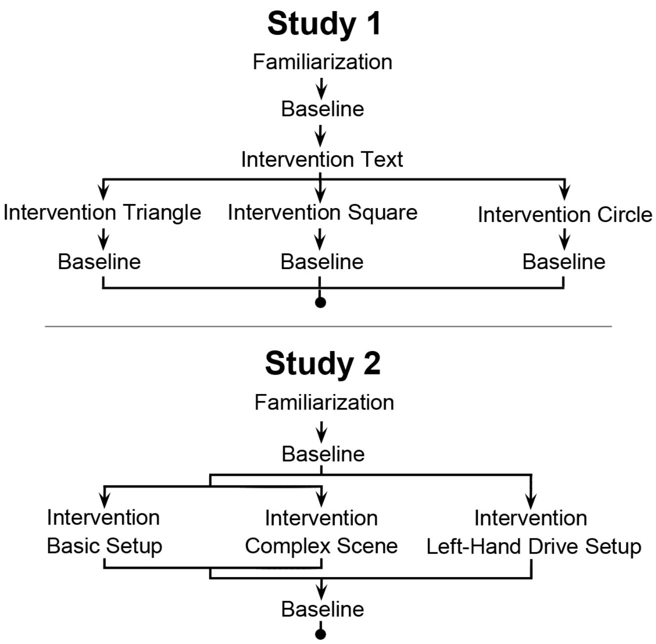

Figure 3. The procedure of the two studies. The order of the parallel tracks was counterbalanced in the studies.

confirm them as invisible (no direct onset visible) and well-visible when looking straight at the vanishing point of the road (0°). Each stimulus was displayed for a maximum of 5 s and disappeared either after this time by itself or earlier in reaction to a button press. After those 5 s and a randomized pause of 3 to 5 s, the next stimulus was displayed at another position.

Procedure

In the beginning, the experimenter introduced the participant to the study's procedure and tasks; the study procedure is depicted in Figure 3. After the participants had performed a 3 min drive to get familiar with the test setup and the driving task, they were asked to fill in a demographic questionnaire. Afterwards, they performed another familiarization track to also get familiar with the WSD and the detection-response task. During this familiarization and all intervention tracks, the participant had to drive (as in the baseline tracks) as well as to respond to the appearing stimuli.

The experimenter then set up the first baseline drive, followed by the first intervention drive which was devoted to textual stimuli. Before the intervention started, the participant was reminded of the priority of the driving task over the secondary task as well as the upcoming target stimulus. When a stimulus appeared, the participant had to detect it, decide whether it was a target stimulus (e.g., female name) or not (then, male name) and if so, press the button on the steering wheel. Then, the stimulus disappeared and after a short break the next stimulus appeared at a different location. If the participant did not respond to a stimulus, it disappeared by itself after reaching its maximum size. After 18 stimuli, the intervention drive was completed. Then, the participant was requested to perform another baseline and three more intervention tracks with triangle, circle, or square as target stimuli and another confounding shape. The overall procedure lasted 35 to 40 min.

Figure 4. The upper scene was used in the first study as well as in the replication and left-hand drive track of the follow-up study. The scene below was used in the follow-up study as a more diverse scene that potentially creates a higher peripheral workload due to higher complexity.

Apparatus

Our test setup is constructed as a wooden frame which holds a real car seat (Figure 1), an acrylic glass pane as a windshield and a gaming steering wheel which controls the driving simulator. The windshield display is realized by three 32" displays positioned below the windshield (aligned on long sides). The images are reflected on the windshield and appear floating 170 cm in front of the driver. The integrated display area measures 2304 × 1360 px and has a field of view of approximately 12.5° to the left, 27.5° to the right and 10° to the top and bottom. The driver's line of sight is 0° (driver looks straight to the vanishing point of the road ahead).

The test setup is calibrated for drivers with a height of 176 cm, according to DIN 33042 [12]; which was suitable for our participants with an average height of 180 cm. The integrated windshield consists of acrylic glass with 140 × 65 cm area and a thickness of 4 mm, positioned at an angle of 39°.

The openDS driving simulator (see top image of Fig. 4) is projected on a wall at a distance of 3.8 m to the driver; the image measures 225 × 177 cm with a resolution of 1920 × 1280 px. We set the vanishing point of the road to be directly in front of the driver (as on a real straight road). The simulator is controlled with a Thrustmaster gaming steering wheel. As the LCT defines a fixed speed of 60 km/h, pedals were available but not used.

Results

Driving Performance

The mean values of the success rate, the mean deviation and the reaction time to the lane change trigger are very constant throughout all driving tracks (see table 1). We performed two separate repeated measure ANOVAs for the mean deviation and the reaction time to the lane change trigger. We did not find evidence for any impacts of the secondary task on the driving performance.

Text	-10°	-5°	0°	5°	10°	15°	20°	25°
7.5°			2089					2017
5°		2136	1988	1846				
0°	1950	1576	1870	1848	1801		2129	1623
-5°		2045	1929	1727				
-7.5°			2181					2401

Circle	-10°	-5°	0°	5°	10°	15°	20°	25°
7.5°			2330					1376
5°		1959	1471	1605				
0°	1476	1484	1174	1325	1345		1470	1365
-5°		1782	1777	1253				
-7.5°			1870					1855

Triangle	-10°	-5°	0°	5°	10°	15°	20°	25°
7.5°			1898					1292
5°		2131	1353	1387				
0°	1476	1220	1319	1140	1076		1276	1377
-5°		1576	1672	1421				
-7.5°			1942					1945

Square	-10°	-5°	0°	5°	10°	15°	20°	25°
7.5°			1730					1303
5°		1818	1307	1420				
0°	1402	1231	1158	1044	1164		1185	1196
-5°		1611	1279	1126				
-7.5°			1400					1594

Figure 5. These patterns show the response times measured during the first study. The intensity of the background color represents the relative duration of the response time. An intense blue refers to low response times. We measured the lowest response times for squares (mean=1340 ms), followed by triangles (mean=1487 ms) and circles (mean=1573 ms). The highest values were obtained for textual stimuli (mean=1946 ms) which we ascribe to the higher complexity. Regarding the positions, we measured the best response times at the level of the horizon but not mandatorily at the center.

Detection-Response Performance

We measured mean response times of 1946 ms (SD=728 ms) for text, 1573 ms (SD=659 ms) for circles, 1487 ms (SD=619 ms) for triangles, and 1340 ms (SD=488 ms) for squares. Further, we calculated the response time for each position and each stimulus type, as shown in Figure 5.

A repeated measure ANOVA with Bonferroni adjusted α-level showed a significant difference for the stimuli type ($p<0.001$, $F(2.42, 38.69)=56.61$, $r=0.78$, Greenhouse-Geisser correction). We obtained the highest values for text, which is not surprising since it is the most complex of the stimuli we tested.

Hence, we compared the shape types again by means of a repeated measure ANOVA with Bonferroni correction and found a significant difference for the shape type ($p=0.001$, $F(1.73, 8.97)=39.8$, Greenhouse-Geisser correction). Post-hoc tests showed a significant difference between the square and circle ($p<0.001$, $r=0.73$) and between square and triangle ($p=0.003$, $r=0.43$) but not between circle and triangle. Consequently, we reject *H1.2* and confirm *H1.3*.

Test Track	Successful LCs	Reaction Time	Mean Deviation
Baselines	16	991 ms	1.41 m
Text	13	981 ms	1.42 m
Triangle	16	984 ms	1.41 m
Circle	15	955 ms	1.37 m
Square	16	935 ms	1.37 m

Table 1. We measured the driving performance as the successfully performed lane changes, the reaction time to the lane change trigger, and the mean deviation between the optimal line and the car's position. The driving performance was nearly identical within all tracks.

As a next step, we looked at the single positions and fields of view. Surprisingly, we found that the response times in the periphery are only slightly higher compared to the center: The response times at an eccentricity of 20 to 25° to the right are lower than the ones measures 10° to the left. These results are not in line with vision research and hence need to be validated with a consecutive study before they can be translated into sizing guidelines. Hence, we do not provide a statistical analysis on the single positions and answer *H1.1* at this point, but will analyze them along with the results of the follow-up study.

Discussion & Implications for Study 2

We measured response times at 17 positions within a field of $35° \times 15°$. We found surprisingly good response times at high eccentricities: The reaction times measured at 20° and 25° in the right part of the windshield are almost equally low as in the central area around the driver's focus point. Further, our results do not support the theory of a symmetric visual perception. These results are not in line with the knowledge about visual perception and hence show need for a follow-up study to validate them before they can be translated into sizing guidelines. We analyzed the study to identify potential reasons for these results in order to control for these reasons and correspondingly design the follow-up study.

At first, we identified the driving simulator as a potential reason. The simulator scene covered the entire WSD area but it did not cover the entire field of regard through the windshield. This might have influenced the visual perception in an unpredictable way. Consequently, the driving simulator scene has to cover the entire windshield area (field of regard) in the second study. Further, we think that the scene (environment in the driving simulator) might have influenced the results; due to its low complexity or brightness or color contrast. So far, we only used one scene with a rural road with very constant surroundings. We decided to test another, more varying and right-hand drive setup & basic scene (replication)

	-10°	-5°	0°	5°	10°	15°	20°	25°
-7.5°			1216					1271
5°		1146	1192	1197				
0°	1409	1340	1122 / 1056	1151	1467		2100	2527
-5°		1333	1376	1210				
-7.5°			1527					2344

complex scene

	-10°	-5°	0°	5°	10°	15°	20°	25°
7.5°			1066					1892
5°		1106	1150	1060				
0°	1254	1121	1178 / 1065	948	1133		1314	1386
-5°		1321	1302	1163				
-7.5°			1648					2180

left-hand drive setup

	-10°	-5°	0°	5°	10°	15°	20°	25°
-7.5°	2382					1467		
5°					1136	1130	1089	
0°	2301	2042		1232	1132	1085 / 993	1136	1139
-5°					1147	1201	1481	
-7.5°	2601					1785		

Figure 6. These patterns show the reaction times measured during the second study. The intensity of the background color represents the length of the reaction time in relation to the minimum and maximum; an intense blue refers to low reaction times. It appears that the reaction times depend strongly on the background and the contrast to it: The basic scene leaded to considerably higher reaction times compared to the complex scene – especially in the periphery. The reaction times measured in the left-hand drive track are very similiar to the ones measured in the right-hand drive track; the major difference is the position P(25°,7.5°) which we ascribe to the background of the scene (bright sky instead of dark mountain).

slightly darker road scene in the second study in order to test for the effects of the scene, its complexity and contrast to the stimuli, on the response times. The simulator scene is still not equivalent to a real driving scene but due to safety and technological issues we are limited to a lab study.

Furthermore, it seems that the lowest reaction times are shifted to the right (from 0°) which would indicate that the participants' gaze was shifted slightly to the right during the study; as in driving curves to the right. Since the lane change directions were counter-balanced and the vanishing point was straight ahead at 0°, we can exclude these factors as potential biases. Another reason for this as well as for the low response times in the periphery might be that the driver's perceptual abilities are not symmetrical. Since the field of regard through the windshield is not symmetrical, the driver might be used to monitoring the right part of the windshield more than (equivalent angles in) the left part. To evaluate this assumption and search for a training effect, we decided to replicate the test in a left-hand drive car.

This leads to three required driving tracks for the follow-up study: (1) a replication of the first study (with full-windshield size driving scene), (2) a replication with a more complex driving scene, and (3) a replication in a left-hand drive car. We reduced the set of stimuli and required a button response to every stimulus to reduce complexity. Also, we decided to not evaluate the driving performance in the second study due to the very constant performance in the first study.

STUDY 2: VERIFICATION UNDER DIVERSE SITUATIONS

The preceding study brought up interesting results which were not expected considering widely accepted literature about visual perception. We conducted a follow-up study to validate these results and also to investigate the reasons for these effects. In particular, we designed three sub-tests according to our assumptions discussed earlier.

Hypotheses:

H2.1: Response time increases with the visual angle of a stimulus (corresponds to *H1.1*).

H2.2: Response time increases in a more complex driving scene compared to the basic scene.

H2.3: Response time increases in a left-hand drive setup compared to a right-hand drive setup.

Independent Variables:

Stimulus Type: Square and circle.

Position: Position at which a stimulus appears.

Dependent Variables (Measurements):

Detection-Response Performance: We evaluate the detection response performance by calculating the response time; the time between the onset of the stimulus and the driver's button press (in milliseconds). The response time can be translated into the stimulus size (in degrees).

Participants

We recruited 22 participants (4 female) by means of social media and e-mail. On average, they were 31 years old (SD=6.6) and 179 cm tall (SD=9 cm). All participants had normal or corrected-to-normal vision (13 of them used glasses or contact lenses) and 21 of them owned a valid driver's license.

Study Design

To ensure the comparability of both studies, we designed the follow-up study similar to the first one. The aim of the follow-up study was to evaluate the results of the first study and investigate potential reasons for our findings. Hence, we decided to conduct three test tracks: (1) replication of the first study, (2) test for the influence of the driving scene (peripheral workload), (3) test for the influence of training (asymmetric visual perception). The order of (1) and (2) is counter-balanced and as a set counter-balanced with (3) in order to reduce the conversion and calibration of the test setup. Participants were assigned randomly to one order.

We applied a within-subjects design with one presented stimulus per position (two stimuli in the center). We applied the LCT and DRT as in the first study but limited this study to two shapes as well as to requesting a button response for both shapes. During each track we displayed 18 shapes to the participants; nine squares and nine circles. To respond to squares, participants had to press a button on the right side of steering wheel; to respond to circles subjects pressed a button on its left side. Thereby, we reduced the complexity of the study design and increased the validity by obtaining larger data sets.

The first test track replicated the first study (*H2.1*). The driver was also seated within a right-hand drive setup and performed

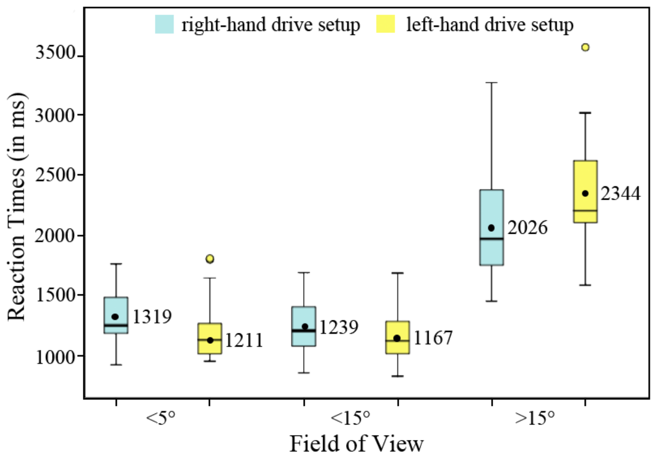

Figure 7. This diagram compares the response times of the left-hand and right-hand drive track for the three fields of view. The response times are slightly lower within the the FoV <5° and <15° but significantly higher at the FoV >15°; though, looking at the positions of this FoV, it appears that this effect has to be ascribed to the position in the sky (P(25, 7,5). Both test tracks used the same driving scene and comparing this position in both setups it becomes very clear that the background of the stimuli caused this difference.

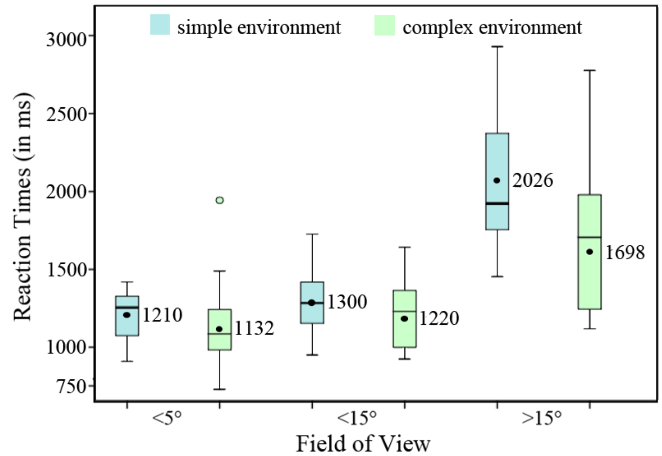

Figure 8. The diagramm depicts the reaction times for the three FoVs measured for the simple and the complex driving scenes. We measured consistently lower reaction times for the complex scene: In the central area around the vanishing point (below 15°) the driving scenes and as follows also the results are very similar. Above 15°, we measured significantly lower reaction times which we ascribe to a better brightness contrast between stimuli and driving scene. The results indicate, that a diverse driving environment leads to a better figure-ground-separation and supports a better stimuli recogntion.

lane changes while responding to the visual stimuli. We ensured that the driving scene covered the entire field of regard through the windshield in order to avoid unwanted and uncontrollable influences, such as low peripheral workload.

The second track tested for the influence of the peripheral workload (*H2.2*). We replicated test track (1) but replaced the driving simulator scene with a more varying one in order to increase the peripheral workload (see Figure 4).

The third track tested for an asymmetric visual perception (*H2.3*). Since our participants were experienced drivers of right-hand drive cars, they are trained to scan their visual field, e.g., for hazards. We assumed this might lead to asymmetric visual perception abilities. To test for such effects, we converted the right-hand drive car setup of test track (1) to a left-hand one; including a re-positioning the steering wheel, the pedals and the vanishing point of the simulated road (to be in front of the driver). Furthermore, we horizontally inverted the stimuli pattern.

Procedure

After introducing the participants to the study procedure and the test setup, the experimenter asked them to fill in the demographic questionnaire. Then, they took a seat in the test setup and adjusted the seat position. The experimenter calibrated the position of the WSD image and introduced the driving task as well as the detection-response task (familiarization drive). The study procedure is depicted in Figure 3.

The test begins and ends with a baseline and comprises three intervention drives in between. During the baseline, the participants have to drive only. During the intervention drive participants have to additionally detect and respond to visual stimuli appearing on the windshield display: The participants had to detect a stimulus that appears on the windshield display and react to it by pressing one out of two buttons (right / left) on the steering wheel, depending on the shape of the stimulus (square / circle). After 3 min of driving and responding to 18 shapes, the intervention drive was completed.

Apparatus

We used the same basic test setup as in the prior study but extended and adjusted it to the requirements of this follow-up study. We optimized the angle of the windshield to 45° and made the seat position and the WSD image adjustable to the individual body dimensions of the participants. We measured and equated the brightness of the three displays and integrated a rail system which allows for a fast conversion of the right-hand drive setup to a left-hand drive setup; including seat, steering wheel and pedals. We also ensured that the driving scene covers the entire field of regard by lining up two projectors (3840 × 1280 px each) next to each other. The projection covers 61° x 30° is directed onto a canvas in 3.5 m distance; leading to an image of 6.3 × 2.0 m.

Results

Detection-Response Performance

We prepared the response times by excluding outliers (values higher than $Q_{3(75\%)}$ + 2.2 × IQR; IQR=interquartile range), missed stimuli and wrongly recognized stimuli; this reduced our data set from 1188 to 1111 trials. Depending on the relation and distribution of the data, we applied Friedman's, Wilcoxon, Mann-Whitney-U, and t-tests with Bonferroni adjusted α-levels. To reduce complexity, we divided the set of positions into three subsets, as visualized in Figure 2: <5°, <15°, and >15°.

The time needed to detect and react to the stimuli at the specific locations are depicted in Figure 6 for each test track. We analyzed the location-specific response times by means of a Friedman test (not normally distributed data). We found a significant influence of the scenario on response time ($\lambda^2(2)$=8.59, p=0.014). We further analyzed this effect by Wilcoxen pairwise comparisons and found significantly different response times for the complex scene (mean=1287 ms, T=26, p=0.007, r=0.58) as well as the left-hand drive setup (mean=1421 ms, T=23, p=0.001 r=0.7) compared to the basic replication of study 1 (mean=1419 ms).

To answer *H2.1*, we compared the FoVs of the basic replication of study 1 by means of t-tests (normally distributed data). We found significantly lower reaction times for <5° compared to <15° (t(21)=-2.82, p=0.01, r=0.27) as well as for <15° compared to >15° (t(21)=-8.35, p<0.001, r=0.77). This shows that the response times generally increase with the eccentricity and confirms *H2.1*.

As a next step, we compared the simple and the complex driving scene regarding the three FoVs <5°, <15° and >15° (see Figure 8). We performed one-sided t-tests (normally distributed data) and found a significant difference for the FoV <15° (t(21)=2.93, p=0.008, r=0.54) and the FoV >15° (t(21)=3.06, p=0.006, r=0.56) but not for the central area (<5°). We expected that the more complex scene leads to a higher peripheral workload and hence to lower response times. Since we found lower response times in the peripheral areas we have to reject *H2.2* for now. The differences between the two scenes might have been to little to actually increase the peripheral workload; a city scene might have resulted in different results. The two driving scenes (see Figure 4) differ in lighting and color but are very similar in the central area around the vanishing point. Looking at the the single positions it appears that we obtained the lower response time always for the scene that provides more contrast to the stimuli; we obtained similar results for the areas where the lighting and color conditions are comparable (e.g., road area).

Equivalently, we analyzed the simple and the converted driving setup regarding the three FoVs <5°, <15° and >15° (see Figure 7). Therefore, we reflected the stimulus pattern of the left-hand drive setup horizontally in order to match the original pattern. We performed a Wilcoxon test (not normally distributed data) and found significantly higher response times for the FoV >15° for the left-hand drive setup compared to right-hand drive setup (T=23, p=0.001 r=0.7). Looking at the single positions of this field of view, though, we only found a considerably different response time for one position (P(25°, 7.5°)). We can ascribe this difference to the change in background color and brightness from dark brown (right-hand drive) to bright sky (left-hand drive). Consequently, despite statistically different results, we do not consider hypothesis *H2.3* confirmed.

Comparison with Preceding Study

To verify the results of the preceding study, we selected the response times by location for circles and squares to match the same stimulus pattern as in the follow-up study. We performed a Mann-Whitney-U test to statistically compare the results of the two independent studies (not normally distributed data). We did not find a significant difference between the overall response times from the two studies (p>0.6). As a next step, we compared the results of the two studies for each

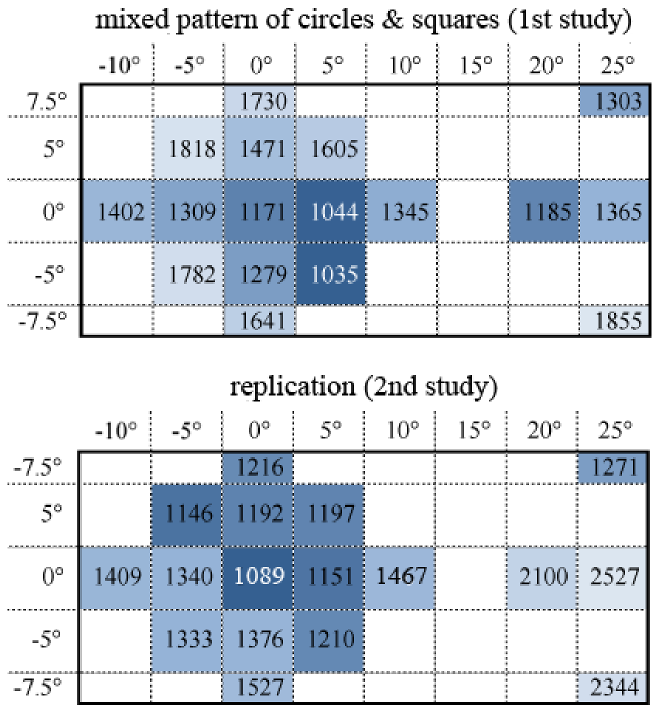

Figure 9. The top graphic merges the reaction times measured for circles and squares during the first study according to the stimulus-position-assignment depicted in Figure 2. We used the same stimulus-position-assignment in the follow-up study. The bottom graphic shows the corresponding reaction times: We measured significantly higher reaction times in the periphery as well as considerably lower reaction times in the upper visual field.

position: In the replication track of the follow-up study, we found significantly higher values at the positions P(20°, 0°) (mean=1185 ms vs. mean=2100 ms; U=24, p<0.001, r=0.66) and P(25°, 0°) (mean=1365 ms vs. mean=2527 ms; U=38, p=0.002, r=0.54); meaning that the response times gathered at these positions during the first study were not confirmed. However, these values are very close to the ones measured with a complex driving scene. We did not find a significant difference for the remaining 15 positions.

High variance in the response times was to be expected due to the high number of variables in the two studies. Hence, differences in the response times should generally not be overrated. Yet, we think that some differences are noteworthy and should be discussed: We measured considerably lower response times for the positions above the horizon P(-5°–5°, 5°–7.5°). Also for the peripheral position above the horizon P(25°, 7.5°), we obtained very good response times compared to the other positions at above 15° eccentricity. The surprisingly low values at P(-5°, -5°) measured in the first study are not confirmed in the replication study. However, it is noteworthy that we found very high response times for the position at P(0°, -7.5°) in every study and condition. Research focusing on finding the optimal position for the small-sized HUD often recommend to center it at P(0°, -5°) and most of the available HUDs do follow this recommendation. Such research does rarely investigate locations below -5°. As mentioned before, we partially assign the differences in the response times at the single positions to normal data variance. Furthermore, variances in the group of

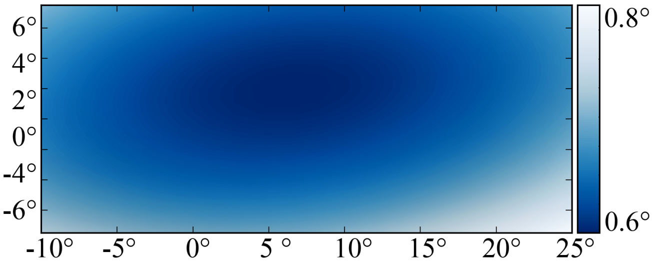

Figure 10. Within the field of regard of 35° × 15°, shapes should be displayed with an angular size of at least 0.6° to 0.8°. The smallest sizes are required to the right of the driver's normal line of sight at around 5°. The recommended stimulus sizes increase towards the periphery in an elliptical shape.

participants or the driving setup might have influenced the results. In particular, we calibrated the WSD for each participant in the replication study and extended the driving simulator scene. This enabled a more precise placement of especially the vertical stimuli and might have resulted in a slightly different placement of the driving scenes. In a consequence, this could influence the saliency of the stimuli and make it easier or harder to detect them in the lab study – however, comparable placement variances come along with a simple change of the environment or adjustment of the seating position or posture.

Sizing Recommendations

The response times measured during both studies can be translated into angular sizes. In order to provide sizing recommendations for the entire tested field of regard (35° × 15°) – also the areas in between the exact tested positions – we interpolated the collected data with a 2D linear regression model. We included quadratic terms into the model, since preceding research indicates that the fields of view are elliptical [14].

We fitted our model on all shape response times (and thus sizes) obtained in the first study, as well as those obtained in the second study, with careful handling of the data from the left-hand driving track: In order to not confound the data of the left and right half of the visual field, we decided to not invert the results obtained below 15° and to exclude the peripheral positions (above 15°) of this track. The resulting fitted model (R^2=0.81) is depicted in Figure 10. Our model allows the smallest sizes (0.6°) to be used in the center with a slight bias to the right, while larger sizes (0.8°) should be used further in the periphery. The area that requires the largest sizes is near the bottom right corner of the display.

We decided to only provide sizing recommendations in this way for shapes (triangle, square and circle), and not for textual stimuli, due to their small data set. However, based on our results, we suggest to display textual stimuli approximately 0.2° larger than shapes.

DISCUSSION & LIMITATIONS

The lane change task is a standardized driving task that requires the driver to perform lane changes on a three lane road. The lack of other road users and the externally triggered lane changes make the task artificial but measurable, as all simulated driving tasks. The lane change maneuver itself is realistic and happens regularly, especially when driving on a highway. A real world study was not possible since head-up displays with a field of regard used in this research are not available yet but also because of safety issues and law restrictions.

The simulated scenes are generally realistic but very constant. As shown in the second study, the color and brightness of the background scene, in particular the contrast to the stimuli, influence the perception of the stimuli. We ensured that all displays are equally bright in order to set equal conditions for all stimuli. We further compared two rural scenes which differ especially in the periphery in brightness and color. The results we obtained in our study are hence valid for the two example scenes but other scenarios, such as driving at night or in a city, might lead to different results: The greater the difference (contrast) between stimulus and background scene, the easier a stimulus can be detected. This is shown in our study by the response times measured for the peripheral positions (>15°) where the contrast differs considerably (except for the lowest position which has asphalt in background) but has already been shown in previous research, e.g., by Strasburger et al. [19]. Further, brightness and contrast depend strongly on the technological realization of the display unit. Real HUDs utilize specific, extremely bright displays or projection systems to ensure the image stays legible in all lighting and weather conditions (which also influence the saliency of the stimuli).

In both studies, we found that the lowest response times are not mandatorily at 0° and rather shifted 5° to the right; this finding is confirmed by Tsimhoni et al. [22]. Generally, we measured lower response times to the right side compared to the left side of the center. Since we controlled for the vanishing point to be directly in front of the driver (0°) and also counterbalanced (and randomized) the direction of the lane changes, the participants' gaze point should (on average) not be shifted to one side due to the study design or apparatus.

Still, a stimulus needs to be of a minimal size to be readable. Our research delivers well-legible sizes for texts and shapes for 17 positions in a field of regard of 35° × 15°. We further derived sizing recommendations for the areas in between those positions by means of a linear regression. Presumably, the minimal stimuli size will further increase towards higher eccentricities but further research is needed to provide exact sizes. Also, we provide sizing guidelines for the tested stimulus types text, circles, triangles, and squares represented as outlines. These stimulus types can be used to represent most of the traffic signs. Other stimuli such as complex icons will most probably require a bigger size to be recognizable. Further, we assume that filled shapes will require a comparable sizing to be legible but potentially a smaller size to be detectable. We assume that stimuli scaled according to our sizing recommendations are legible in a real world scenario given a comparable or even better contrast ratio between display and environment. Yet, due to the safety risk of unreadable information, we think that the sizing guidelines should be evaluated in a real world study once an appropriate windshield display prototype exists.

CONCLUSIONS & FUTURE WORK

We performed two user studies on the recognition of simple stimuli – text, circles, triangles, and squares – presented on a large-sized head-up display. We measured the response times and derived stimuli sizes for 17 positions within a left-hand and a right-hand drive setup and within two different simulated environments.

Our results show that the response times for the four stimulus types differ: Squares can be recognized at the smallest size, followed by triangles and circles. Text needs considerably larger sizes, presumably due the its higher complexity. Overall, response times increase along with the eccentricity of the stimuli; this is in line with basic vision research. Also, response times increase for lower contrast between stimuli and background. This becomes especially clear when comparing the results obtained for the two different scenes in the second study. Based on our results, we present recommendations for minimal stimuli sizes for a field of regard of $35° \times 15°$. Our sizing guidelines help HUD designers to provide interfaces which are well-legible and avoid clutter in the driver's view and the occlusion the road scene at the same time.

In contrast to existing research, the presented studies investigated a considerably larger field of regard and focused on response times to derive the stimuli sizes instead of the optimal HUD position. Yet, future research should aim to investigate stimuli sizes for a even larger field of regard – preferably the entire windshield area. Further, the effect size of the color and brightness contrast between stimuli and background scene and its influence on the recognizable stimuli sizes needs to be understood better. Therefore, a follow-up study should investigate other, e.g. urban, environments as well as different daytime and weather conditions. Once head-up displays can display content at different depth levels and using augmented reality, stimuli that move, especially an (greater) depth, should be investigated – especially since the size and depth perception interrelate strongly.

ACKNOWLEDGMENTS

We thank the company IAV GmbH for the funding and the support.

REFERENCES

1. Miyoshi Ayama and Masato Sakurai. 2003. Changes in hue and saturation of chromatic lights presented in the peripheral visual field. *Color Research & Application* 28, 6 (2003), 413–424. DOI: http://dx.doi.org/10.1002/col.10194
2. A. S. Cohen and R. Hirsig. 1990. Zur Bedeutung des fovealen Sehens für die Informationsaufnahme bei hoher Beanspruchung. (1990), 47–58.
3. D. Crundall, G. Underwood, and P. Chapman. 1075-1087. Driving Experience and the Functional Field of View. 28(9) (1075-1087), 1999.
4. David Crundall, Geoffrey Underwood, and Peter Chapman. 1999. Driving experience and the functional field of view. *Perception-London* 28, 9 (1999), 1075–1088.
5. Ronald Ecker. 2013. Der verteilte Fahrerinteraktionsraum. (2013).
6. K. W. Gish and L. Staplin. 1995. Human Factors Aspects of Using Head Up Displays in Automobiles: A Review of the Literature. (1995).
7. Renate Haeuslschmid, Bastian Pfleging, and Florian Alt. 2016. A Design Space to Support the Development of Windshield Applications for the Car. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA.
8. Renate Häuslschmid, Laura Schnurr, Julie Wagner, and Andreas Butz. 2015. Contact-analog Warnings on Windshield Displays Promote Monitoring the Road Scene. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. ACM, New York, NY, USA, 64–71. DOI: http://dx.doi.org/10.1145/2799250.2799274
9. M. J. M. Houtmans and A. F Sanders. 1984. Perception of Signals Presented in the Periphery of the Visual Field. *Acta Psychologica* 55, 2 (1984), 143–155. DOI: http://dx.doi.org/10.1016/0001-6918(84)90064-7
10. British Standards Institution. 2010. Road vehicles – Ergonomic aspects of transport information and control systems – Simulated lane change test to assess in-vehicle secondary task demand (ISO 26022:2010). (2010).
11. Arihiro Isomura, Koichi Kamiya, and Katsunori Hamatani. 1993. Driver's Cognition in Peripheral Field of View. (1993). DOI: http://dx.doi.org/10.4271/931876
12. Hans W. Juergens. 2004. Erhebung anthropometrischer Maße zur Aktualisierung der DIN 33 402 Teil 2. (2004).
13. N. Kishishita, K. Kiyokawa, J. Orlosky, T. Mashita, H. Takemura, and E. Kruijff. 2014. Analysing the Effects of a Wide Field of View Augmented Reality Display on Search Performance in Divided Attention Tasks. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 177–186. DOI: http://dx.doi.org/10.1109/ISMAR.2014.6948425
14. Dave Lamble, Matti Laakso, and Heikki Summala. 1999. Detection Thresholds in Car Following Situations and Peripheral Vision: Implications for Positioning of Visually Demanding in-car Displays. *Ergonomics* 42, 6 (1999), 807–815.
15. T. Lino, T. Otsuka, and Y. Suzuki. 1988. Development of Heads-Up Display for a Motor Vehicle. (1988), 15–23. DOI: http://dx.doi.org/doi:10.4271/880217
16. L. R. Newsome. 1972. Visual angle and apparent size of objects in peripheral vision. *Perception & Psychophysics* 12, 3 (1972), 300–304. DOI: http://dx.doi.org/10.3758/BF03207209
17. Tony Matthias Poitschke. 2011. Blickbasierte Mensch-Maschine Interaktion im Automobil. (2011).
18. Thomas A. Ranney, G. H. Scott Baldwin, Larry A. Smith, Elizabeth N. Mazzae, and Russell S. Pierce. 2014. Detection Response Task (DRT) Evaluation for Driver Distraction Measurement Application. National Highway Traffic Safety Administration (NHTSA) (report no: DOT HS 812 077).

19. Hans Strasburger, Lewis O. Harvey, and Ingo Rentschler. 1991. Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics* 49, 6 (1991), 495–508. DOI: http://dx.doi.org/10.3758/BF03212183

20. Heikki Summala, Tapio Nieminen, and Maaret Punto. 1996. Maintaining lane position with peripheral vision during in-vehicle tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 38, 3 (1996), 442–451.

21. Omer Tsimhoni. 2000. *Display of short text messages on automotive HUDs: Effects of driving workload and message location*. Technical Report. http://www.researchgate.net/profile/Omer_Tsimhoni/publication/30817801_Display_of_short_text_messages_on_automotive_HUDs_effects_of_driving_workload_and_message_location/links/0c960519f0b7ce54b4000000.pdf

22. Omer Tsimhoni, Paul Green, and Hiroshi Watanabe. 2001. Detecting and reading text on HUDs: effects of driving workload and message location. In *11th Annual ITS America Meeting, Miami, FL, CD-ROM*. http://www.researchgate.net/profile/Omer_Tsimhoni/publication/245895338_Detecting_and_Reading_Text_on_HUDs_Effects_of_Driving_Workload_and_Message_Location/links/00b4952d7e8f55736a000000.pdf

23. Julia Werneke and Mark Vollrath. 2011. Signal evaluation environment: a new method for the design of peripheral in-vehicle warning signals. *Behavior Research Methods* 43, 2 (2011), 537–547. DOI: http://dx.doi.org/10.3758/s13428-010-0054-8