Comparison of an old and a new Head-Up Display design concept for urban driving

Martin Götze, Christian Schweiger, Johannes Eisner, & Klaus Bengler
Institute of Ergonomics, Technische Universität München
Germany

Abstract

In the future, full use of advanced driving assistance systems will move from highways and freeways to urban areas. This additional assistance systems use case may require the communication of additional information and warnings. Consequently, how information is currently presented in head-up displays will need to change in order to not visually overload the driver. Current head-up design concepts are usually arranged in three clusters: the driven speed, navigation information, and various assistance systems. The new design concept needs to be more generic, driver-focused, and action-orientated. For this experiment, an old design concept and a potential new one were compared in terms of reaction time, response accuracy, and a subjective evaluation. Thirty participants (N = 30) performed an occlusion task and a choice reaction time task to find possible differences between the designs. Statistical tests were performed to examine global and specific accuracy. The occlusion task showed no significant difference between the two designs. However, the new design yielded better CRT performance in terms of response accuracy and reaction times. Then again, the subjective measures showed an advantage of the old HUD design. In conclusion, this paper shows the benefits and downsides of a new urban HMI concept.

Introduction

Driving a vehicle is a visual and cognitive burden for the driver, especially in urban areas (Recarte & Nunes, 2003; Rumar, 1990). Nevertheless, looking at the number of accidents in recent years the number of road fatalities has fallen sharply (Statistisches Bundesamt Wiesbaden, 2014). First and foremost, with 93%, is human error which has made a major contribution (Winner et al., 2012). Analysing the types of human error involving personal injury, it can be seen that most of them could be avoided, or at least assisted, by a suitable human-machine-interface (HMI) design providing the right information in the right way. For this purpose, a generic and integrative HMI design concept for head-up displays (HUD) is proposed, with warnings and information presented in a driver-focused and action-orientated manner (Petermann-Stock & Rhede, 2013).
The scenarios and characteristics of urban areas are quite different from rural areas or highways with a higher frequency of necessary manoeuvres (Gevatter & Grünhaupt, 2006). In addition, the city environment is characterised by a much higher complexity (Schröder, 2012) and multiple road users like other cars, trains, bikes and even weaker traffic participants (Winner et al., 2012), see Figure 1. Thus, about 68% of all accidents in 2010 involving personal injury occurred in these areas. Some of the reasons were also seen in the increased decision making, the much shorter time window and the fast sequence of traffic notifications or road signs (Götze et al., 2014).

Figure 1. Typical urban area scenarios with multiple road user and a much higher complexity compared to rural areas. © iStock.com/Bim

The head-up display is well known from aviation and was first introduced in an automotive context by Bubb (1978). In comparison with head-down displays, like the instrument cluster or any other in-vehicle displays, the main advantage is the virtual image (2 - 2.5 m in front) in the line of sight of the driver (see Figure 2), which leads to significantly faster reaction times to the presented information (Wittmann et al., 2006) and still about 40% - 50% visual acuity on the road or driving scene (Schmidtke & Bernotat, 1993). In addition, a significantly lower workload presenting the same warnings as in a head-down display has previously been found (Götze & Bengler, 2015). Nonetheless, there is a risk of overloading the display with too much unnecessary information, especially in urban areas where this new display concept approach is presented and compared to presently available solutions.
head-up display concepts for urban driving

Method

One part of the experiment was conducted using the occlusion technique (ISO 16673:2007). Several studies have showed the validity, reliability and suitability of this method for evaluating or comparing different in-vehicle display design concepts (Baumann et al., 2004; Gelau et al., 2009; Horst, 2004; Noy et al., 2004). Occlusion is classified as a universal tool for measuring the time and accuracy when reading information on HMI displays. The method is used to assess the visual demands of the tasks. The independent variables for this tool are the time period in which a certain stimulus is presented and the time frame between two of those stimuli (Baumann et al., 2004). This study was done by limiting the presentation time of the stimulus on the screen.

The other part of the experiment consisted of two parallel tasks; a choice reaction time task (CRT) and a continuous tracking task (CTT). The aim of this experimental part was to observe the head-up display and to respond to different stimuli using a keypad, while simultaneously doing the CTT on a projection screen. The CRT technique is used to measure reaction times and accuracy when responding to presented stimuli. The method was applied in such a way that the stimulus disappeared after 1500 ms, or after the participant pressed a button to indicate whether or not the answer was correct. Additionally, the time between the stimuli varied between 3000, 5000 and 7000 ms. The CTT is a visual-manual task requiring continuous control by the participants (Eichinger, 2011). The task in this type of CTT was to control the position of a vertically and horizontally moving cross-hair towards a central point using a joystick. The task was used to represent and simulate the driving task and was therefore presented on a projection screen in front of the vehicle.

Figure 2. Comparison of different angles and display distances between the head-up and head-down display (Milicic, 2010).
Figure 3. Different current state head-up display designs divided into three clusters; a less detailed version left from Breisinger (2007) and a more detailed one from a Kia K9 Quoris.

Stimuli

Current state head-up display design concepts were researched and categorised. As a result, it was found that the majority are divided into three different clusters. One cluster showed the current speed combined with traffic sign recognition, one displayed navigation information, and the last one provided advanced driving assistance systems information. The order of the three clusters might have differed from brand to brand, but the content was still the same. The one from Breisinger (2007) in Figure 3 was used as an “old” design pattern and is further referred to as “cluster design”. The solution for a “new” HUD design concept might be a generic and integrative approach, presenting the relevant information centred and packed so that the driver can gather all the important information at a first glance without first needing to scan the whole head-up display.

Figure 4 shows examples of both of the compared concepts with the categories of speed, navigation, and warning. The question raised for the “speed” category was always about the driven speed being below or above the shown speed limit. The second category, “navigation”, was used to ask about the direction of the navigation arrow or the distance to the next turn which resulted in two different targets at the design concept. The last category, “warning”, showed two different warning signs (small and large). Not all font and symbol sizes match in both design concepts due to the different approaches of presenting the information. Nevertheless, all the font sizes met the requirements of ISO 15008 (2009) which recommends a minimum size of 0.2°. As an additional measure, a coloured and monochrome version were used for the generic design to examine the influence of colour on reaction times and accuracy.
Framework Conditions of the Study

The study took place in a stationary 2008 BMW X5 (E70) with a pre-installed series head-up display. The resolution of the HUD was 480x240 px (approx. 20 x 10 cm) and it was able to display the colours red, yellow and orange. In the vehicle, a joystick (right-hand) and a keypad (left-hand) were installed to perform the CRT and CTT. The vehicle had been parked and centred 4.50 m in front of a projection screen (see Figure 5). The distance complied with the 85th percentile of the length of compact-class vehicles from 2010 (Schuster et al., 2011) and simulated the fixation of the brake lights one vehicle length ahead. The light level was always kept in the scotopic range.

The occlusion task consisted of three different stimuli (navigation, speed, and distance) each with four variations. In addition, three cycles were performed resulting in 36 stimuli for each HUD design concept (= 72 stimuli each participant).
For the CRT task, a warning stimuli was added and the distance stimuli removed to fit with the used keypad. Altogether, three different stimuli were included (navigation, speed, and warning symbols) with four variations again for the first two stimuli, along with two variations for the warning symbols. Participants had to do eight cycles of the ten stimuli with an overall 80 CRT stimuli for each HMI design concept (= 160 stimuli each participant).

**Procedure**

When the participants arrived, all of them performed a visual acuity test (Landolt ring test) to ensure that each participant met the minimum acuity for driving (0.5 according to Colenbrander and De Laey [2005]). This was especially important considering the nature of the experiment in which two design concepts are compared in terms of reading information. Immediately after, participants filled out a demographical questionnaire and were introduced to the experiment and its objective. One of the experimental parts then started in a permuted order. The individual instructions for each part were given at the particular time the task started, always followed by a training session to practice each task. Additionally, the Post Study System Usability Questionnaire (PSSUQ) was filled out during the CRT part. At the end, each participant answered a final questionnaire, rating each concept.

For the occlusion task (see Figure 6), participants read the instructions and question (italic) for each stimuli on the projection screen. It was shown for 3000 ms. Immediately after, a fixation cross appeared for 2000 ms to ensure the eyes were focused on the screen and not yet on the HUD. Finally, the occlusion stimuli was presented in the head-up display for either 200, 250, or 300 ms followed by the question again (bold) until the response was given. Only the accuracy was measured.

![Figure 6. Procedure and duration of the occlusion task on each display.](image)

During the CRT task, participants operated the joystick and simultaneously, stimuli were presented on the HUD for 1500 ms or until a button was pressed. The keypad consisted of four buttons: left, right, up, and down. Participants answered the speed stimuli (to high ↑, to low ↓), the navigation stimuli (arrow left ←, arrow right →), and the warning stimuli (either key is correct).
The PSSUQ is a questionnaire used to measure the usability of a system (Lewis, 2002). The PSSUQ can be classified as a psychometric tool to initially evaluate computer systems. It was developed by IBM (Lewis, 1992) and has a reliability of 0.83-0.96. To match the questionnaire with this study, the phrase “system” was replaced by “display concept” to avoid flawed ratings of different systems.

All displayed questions, instructions or stimuli in this study were prepared and executed with E-Prime 2.0 (Psychology Software Tools, Inc.). The performance metrics (accuracy and reaction time) were also recorded with E-Prime.

**Results**

Thirty-one healthy volunteers participated in this study (11 women, 20 men). The participants had between 18 and 52 years of age (M = 27.1, SD = 9.6). All of the participants passed the visus test (M = 0.9 visus). One of the participants was excluded from the occlusion task due to technical problems.

**Occlusion Task**

The mean global accuracy was calculated for 30 participants and all three time-frame conditions. No response, or reporting an inaccurate answer, were considered as errors. A paired-sampled t-test was executed to examine any difference in global accuracy rate. The mean and SDs for each design concept can be found in Table 1. No significant difference was found.

<table>
<thead>
<tr>
<th></th>
<th>Cluster Design</th>
<th>Generic Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>77.7 %</td>
<td>78.1 %</td>
</tr>
<tr>
<td>SD</td>
<td>9.9 %</td>
<td>10.2 %</td>
</tr>
</tbody>
</table>

The mean accuracy for all three different presentation times and both design concepts was calculated. A one-way repeated measure ANOVA was conducted. Mauchly’s test indicated that the assumption of sphericity had not been violated. The results show that there was no significant effect. These results suggest that the design of the display concept has no effect on the accuracy of reading and reporting the information from the HUD. Accuracy percentages and standard deviations are given in Figure 7. In addition, there was a significant main effect of presentation time, $F(2, 28) = 10.945$, $p \leq .001$, suggesting that presentation time significantly affected mean accuracy. This effect is well-known from previous studies (Götze et al., 2013).
The mean global accuracy was calculated for all 31 participants and all three stimuli categories. Only correct reaction times were calculated. A paired-sampled t-test was executed to examine any difference in global accuracy rate. The mean and SDs for each design concept can be found in Table 2. Again, no significant difference was found.

Table 2. Mean global response accuracy for two design concepts using a CRT task (N = 31).

<table>
<thead>
<tr>
<th></th>
<th>Cluster Design</th>
<th>Generic Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>89.24 %</td>
<td>91.37 %</td>
</tr>
<tr>
<td>SD</td>
<td>7.1 %</td>
<td>5.5 %</td>
</tr>
</tbody>
</table>

The reaction times were calculated for all participants and all three stimuli categories. A paired-sampled t-test was executed to examine any difference in global mean reaction time. The mean and SDs for each design concept can be found in Table 3. A significant difference was found between the two designs; t(30) = 6.5071, p ≤ .001.

Table 3. Mean global reaction times for two design concepts using a CRT task (N = 31).

<table>
<thead>
<tr>
<th></th>
<th>Cluster Design</th>
<th>Generic Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>954.9 ms</td>
<td>907.9 ms</td>
</tr>
<tr>
<td>SD</td>
<td>69.7 ms</td>
<td>64.9 ms</td>
</tr>
</tbody>
</table>

Additionally, all the accuracies were individually calculated for the three different stimuli categories. A one-way repeated measure ANOVA was conducted. Mauchly’s test indicated that the assumption of sphericity had been violated, X²(2) = 6.591, p =0.037; the degrees of freedom were therefore corrected using Greenhouse-Geisser estimates of sphericity (ε = 0.831). The results show that there was no significant effect of the design concept on mean accuracy for the different stimuli categories. The means and standard deviation can be found in Table 4. These results suggest,
again, that the design of the concept has no effect on the accuracy of reading and reacting to information from a head-up display.

Table 4. Mean accuracy (SD) and mean reaction times (SD) for all three categories (N = 31).

<table>
<thead>
<tr>
<th></th>
<th>Speed</th>
<th>Navigation</th>
<th>Warning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster Acc</td>
<td>80.1 % (14.7)</td>
<td>95.3 % (6.5)</td>
<td>95.8 % (7.2)</td>
</tr>
<tr>
<td>Generic Acc</td>
<td>82.3 % (12.0)</td>
<td>96.2 % (4.2)</td>
<td>98.4 % (6.8)</td>
</tr>
</tbody>
</table>

Furthermore, all the reaction times were individually calculated for the three different stimuli categories. Here, again, a one-way repeated measure ANOVA was conducted. Mauchly’s test indicated that the assumption of sphericity had not been violated. The test found a significant effect of the design on mean reaction times for the three different categories $F(1, 30) = 45.495, p \leq .001$.

This finding made it necessary to execute paired-sampled t-tests comparing each design concept for each category. The results were corrected by the Holm-Bonferroni Sequential Correction (Holm, 1979). All means and the standard deviations are shown in Figure 8. A significant difference was found when comparing the two design concepts in the speed category; $t(30) = 4.278, p \leq .001$. In addition, a significant difference was found for the warning category; $t(30) = 6.189, p \leq .001$. No significant difference was found for the navigation task. These findings suggest that the generic design concept is beneficial in terms of reaction times compared to the cluster concept.

Figure 8. Mean reaction times with standard deviation for all three categories (N = 31).

Questionnaires

For the PSSUQ, the mean of all 19 items was calculated. The overall score for both design concepts can be found in Table 5. A lower score implies a better rating. Executing a pair-sampled t-test found no significant differences.
The final questionnaire included a question about the satisfaction with each design concept, rating it on a seven-point Likert scale. The question was scaled from “I do fully agree” (1) to “I do not agree at all” (7) and therefore a lower score being a better rating. The cluster design was rated 2.77 (1.6) and the generic design with 3.13 (1.7). No significant difference was found.

Discussion

The main aim of this study was to evaluate a new approach to a head-up display design concept for urban driving with a current state version. No difference in terms of colour was found. Looking at the occlusion results, no significant difference was found either in respect to the global accuracy, or for the specific accuracies of the different presentation times. Based on those results, neither of the two design concepts are superior to each other. However, looking at the choice reaction time task, a significant difference in the global reaction time, as well as in the individual reaction time in the three categories has been seen. In detail, responses to the navigation question and to the warning signals resulted in a significantly better result for the generic version. This is particularly interesting, since the font and traffic sign size is smaller in the generic version. Despite that, again, no difference has been found in all three categories in respect to the accuracy between the two design concepts. Participants were equally accurate, but with different reaction times.

Conclusion

Neither design concept seems to be ultimately superior when compared with each other. Still, both HUD designs do show benefits in individual places. Further research needs to be done to assist the driver with urban driving in the best possible way. Moreover, additional ADAS need to be added and considered when evaluating the concept. Furthermore, in order to follow the integrative approach (Götzte et al., 2015), the adding of more in-vehicle components is necessary which might then result in different design options.

In conclusion, the study shows the potential of a new generic design approach for head-up displays in urban driving. While the reaction times were significantly shorter with the new concept compared to the cluster design, the subjective rating showed some contrary effects. In a future study, the benefits of both design concepts will be considered to build an even more suitable approach to urban driving.
References


