3D Displays in Cars: Exploring the User Performance for a Stereoscopic Instrument Cluster

Nora Broy[1,2,3], Florian Alt[1], Stefan Schneegass[3], Bastian Pfleging[3]

[1]University of Munich
Media Informatics Group
Amalienstraße 17
80333 Munich, Germany
firstname.lastname@ifi.lmu.de

[2]BMW Group
Research & Technology
Hanauer Straße 46
80992 Munich, Germany
Nora.NB.Broy@bmw.de

[3]University of Stuttgart
VIS
Pfaffenwaldring 5a
70569 Stuttgart, Germany
firstname.lastname@vis.uni-stuttgart.de

ABSTRACT

In this paper, we investigate user performance for stereoscopic automotive user interfaces (UI). Our work is motivated by the fact that stereoscopic displays are about to find their way into cars. Such a safety-critical application area creates an inherent need to understand how the use of stereoscopic 3D visualizations impacts user performance. We conducted a comprehensive study with 56 participants to investigate the impact of a 3D instrument cluster (IC) on primary and secondary task performance. We investigated different visualizations (2D and 3D) and complexities (low vs. high amount of details) of the IC as well as two 3D display technologies (shutter vs. autostereoscopy). As secondary tasks the participants judged spatial relations between UI elements (expected events) and reacted on pop-up instructions (unexpected events) in the IC. The results show that stereoscopy increases accuracy for expected events, decreases task completion times for unexpected tasks, and increases the attractiveness of the interface. Furthermore, we found a significant influence of the used technology, indicating that secondary task performance improves for shutter displays.

Author Keywords

Automotive UIs; stereoscopic 3D; user performance

ACM Classification Keywords

H.5.2 **[Information Interfaces and Presentation]**: User Interfaces—*Screen design (e.g., text, graphics, color)*

INTRODUCTION

Stereoscopic 3D (S3D) displays are quickly proliferating in our everyday life. Having been a technology mainly used in the entertainment sector for the past years, the advent of commercial autostereoscopic devices may soon make this technology ubiquitous. One reason for this is that with autostereoscopy users can perceive 3D content without having to wear glasses. Particularly, the automotive industry takes notice of this development. For example, Mercedes integrated a stereoscopic instrument cluster into their concept car F125[1]. This is just one example for a paradigm shift, where analogue displays are increasingly replaced by their digital counterparts.

Figure 1: Setup of our experiment: We use a driving simulator to evaluate the influence of visualization, technology, and interface complexity on user performance and subjective perception of S3D user interfaces.

Our work is motivated by the fact that 3D displays provide novel means to display information. Particularly, when presented in an abstract form (e.g., icons, warnings), information that relates to each other can be spatially grouped. It can be positioned in a way that allows users to quickly and accurately perceive the priority of UI elements (i.e., more important information in the front). Furthermore, the third dimension can be exploited to communicate the spatial relationships between two items – the distance between two items shows how far two events are apart, either time-wise or location-wise. As the driver is mainly engaged in the primary task of driving a car such an interface may be useful to communicate information in a subtle and unobtrusive way. For example, a

[1]`http://www.pocket-lint.com/news/112047-mercedes-benz-f152-concept-car-video`

AutomotiveUI'14, September 17 - 19 2014, Seattle, WA, US
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3212-5/14/09 ... $15.00.
http://dx.doi.org/10.1145/2667317.2667319

navigation system can communicate the distance to the next exit on the current highway based on depth information. So far it is unclear how the use of S3D UIs impacts the user.

We investigate how S3D UIs influence user performance. This is crucial in cars, since the driver's attention towards the display needs to be minimized. We present a driving simulator study which evaluates the visual and cognitive load implied by the use of 3D displays. We focus on primary (driving) and secondary tasks to reflect different attention levels.

First, we created a 3D digital instrument cluster (IC). The design is based on concrete guidelines we extracted from relevant literature. The IC is able to communicate various types of information. Besides driving-relevant information, such as the current speed, the instrument cluster provides traffic information, warnings, distances, and driving instructions. Second, we conducted a user study in a standardized driving scenario where users needed to respond to expected and unexpected events. We compare a 2D against a 3D visualization, interfaces of different complexity (low / high amount of information), and different S3D technologies (shutter vs. autostereoscopy). For each participant we measure primary and secondary task performance, as well as subjective perception of the UI (simulator sickness, workload, attractiveness).

The results show that 3D is well suited for non-primary tasks in the car. We did not only find an increase in accuracy for judging the distance to a UI object (for expected events), but also that task completion time decreases for responding to unexpected events. The use of 3D does not significantly influence driving performance, compared to the standard 2D UI. But there is a significant influence of the display technology. Finally, subjective feedback reveals that participants favor the S3D visualization of the UI over its 2D version.

The contribution of this work is twofold: First, we present a driving simulator study, comparing how different aspects of S3D UIs influence performance and perception during different tasks. Second, we discuss core design aspects with the aim to help designers create future S3D UIs.

BACKGROUND AND RELATED WORK

In this section we provide an overview of related work and background on 3D display technology, 3D human factors (e.g., cognitive load), 3D automotive UIs, and automotive UIs testing procedures to differentiate and motivate our work.

3D Display Technology

3D display technologies create a 3D effect on a 2D screen by presenting different images for each of the viewer's eyes [29]. The slight horizontal difference between left and right eye image that evokes the 3D impression is called binocular disparity or parallax. Some 3D technologies require glasses to create the different images (e.g., shutter). However, wearing such glasses while driving a car can be bothersome since they may distract from the driving task, for instance, by darkening the user‘s view. In contrast, (glasses-free) autostereoscopic displays have the requirement that the user is at a specific position to create an optimal 3D effect. This is challenging for most settings but easily achievable in an in-car setting.

S3D and Human Factors

3D displays have some benefits compared to their monoscopic counterparts. Studies show that S3D increases user experience (UX). For instance, Schild et al. show that S3D increases the UX for games [26] while Häkkilä et al. discover similar effects for mobile phones [11]. Moreover, highlighting objects using S3D decreases visual search times [13]. A summary of the benefits of stereoscopic displays is given by McIntire et al. [20]. The additional third dimension can be used to further group information onto depth layers (*layered 3D UIs*). Layered 3D UIs can increase user performance [21] if the parameters of such UIs are chosen wisely [4].

Van Beurden et al. discuss the influence of 3D on cognitive load [30]. They compare monoscopic and stereoscopic displays showing that S3D reduces the workload. Wittmann et al. show the same effect for air-traffic controllers [32]. Broy et al. explore the applicability of 3D displays in the automotive context [5]. They found no differences between 2D and 3D displays regarding cognitive load.

Previous work show that 3D displays can cause visual fatigue and discomfort [18]. Main reasons are the use of excessive parallaxes and accommodation-convergence mismatches (for 3D displays viewers focus on the screen while converging on the object in 3D space). The interface designer needs to follow specific guidelines to create usable layered 3D UIs [4].

3D Visualizations in automotive UIs

First experiments for 3D visualizations in automotive UIs show that monoscopic 3D visualizations are preferred by the users (better usability) and that they reduce the task completion time for short tasks compared to 2D list-based UIs [6]. Regarding autostereoscopic 3D displays, Krüger found that such displays require longer attention spans and they were not considered being more attractive than traditional displays for the use case of adaptive cruise control (ACC) [17]. However, high quality stereoscopic visualizations were shown to support prioritizing the foremost content and to increase the perceived quality of an in-vehicle information system [5].

Measuring Cognitive Load

We focus on cognitive load as one aspect of workload. Different methods allow the assessment of the user's cognitive load. Subjective and objective methods are distinguished:

Subjective Methods. These methods rely on the user's estimation of their cognitive load. Such information is often collected through standardized questionnaires. While self-ratings may appear questionable, Gopher et al. show that self-assessment can provide reliable insights into cognitive load [9]. A common questionnaire is the NASA Task Load Index (NASA TLX) [12], that uses rating scales to assess the user's workload. It was adopted for automotive settings, e.g., the Driver Activity Load Index (DALI) [24].

Objective Methods. These methods either measure user performance or the physiological user condition. Measures that correlate with workload include *physiological measures* such as heart rate [25] and skin conductance [8]. Van Orden et al. [31] investigated eye movements and showed that blink, fixation frequency, and pupil diameter correlate with cognitive load. For *performance-based measures* two subclasses can be distinguished [23]: First, the user performance for the primary task can be measured. Popular metrics include task completion time or detection rates [7]. Commonly used tasks include solving mathematical problems [22]. Second, the performance of a secondary, concurrently performed task, can be measured, e.g., the peripheral detection task [15].

To provide a comprehensive assessment of cognitive load we use subjective as well as objective methods in our work.

User Performance in a Driving Simulator

Since the driver always needs to share attention between the primary task of driving the car and secondary tasks (e.g., interacting with in-vehicle systems), additional measures to control the primary task performance need to be taken into account. This comprises information about lateral control (e.g., lane deviation, steering wheel activity), longitudinal control (e.g., maintaining speed, braking behavior), and driver reaction (e.g., recognition time for unexpected incidents) [3].

The evaluation of an automotive UI therefore often combines measuring the task performance of a secondary task as well as its influence on the primary task performance. Such evaluations can be done either with abstract primary tasks in a lab (e.g., occlusion tests to measure visual demand [14]), in driving simulator setups (e.g., Lane Change Task (LCT) [19]), or in real-world experiments. We focus on simulator studies as these allow for a detailed control of environmental conditions.

We chose the following headway test as proposed in the AAM Guidelines [1] as this test is supposed to provide a good tradeoff between a realistic driving situation and a controlled environment. The idea is to test the impact of solving a secondary (interaction) task on the primary driving task by directly assessing concurrent driving performance under dynamic conditions in a simulated environment. The performance is then compared to the performance of accepted reference tasks. As a primary task the driver shall follow a lead vehicle while maintaining a constant distance to this vehicle.

HYPOTHESES

We focus on three main hypotheses that concern the influence of visualization, technology, and interface complexity on primary and secondary task performance, gaze behavior, and subjective measures, including UX, driver activity load, and simulator sickness.

H1: Influence of 3D Visualization. First, we explore differences inferred by the 3D visualization chosen for presenting the UI. Usually, UI designers need to make a choice whether to create a monoscopic (2D) or a stereoscopic (3D) visualization – independent of how it is later presented technically. We hypothesize that a 3D visualization has a positive effect on secondary task performance as well as on UX. At the same time, we expect an influence on the primary task performance, the time users take their eyes off the road and on the driver activity load. However, it is uncertain if this influence is positive or negative. On one hand, 3D displays may support the driver in the primary driving task by making relevant information easier perceivable and, thus, decrease the load for assessing the information. On the other hand, accommodation-convergence mismatches and visual artifacts, can affect the cognitive load, the driver's condition and hence the primary driving task.

H2: Influence of Technology. Second, we investigate the influence of the used display technology. Whereas we consider a 2D screen, today commonly found for digital dashboards, as a baseline, we are particularly interested in comparing two S3D technologies: glass-based technologies (i.e., active shutter) and glasses-free technologies (i.e., autostereoscopic displays). We hypothesize autostereoscopy to decrease UX and secondary task performance since autostereoscopic technologies lack in 3D quality compared to shutter. In contrast, we think that shutter has a negative impact on the driver activity load and the primary task performance since the required glasses are disturbing and darken the view.

H3: Influence of Complexity. Third, we investigate the complexity of the UI. Whereas in traditional dashboards most elements are hard-wired which limited the way in which the available space could be used, digital displays allow for displaying any information available in the car. Hence, designers of future car UIs may be intrigued to display as much information as possible. However, we hypothesize that this overload leads to a decrease in the UX, primary and secondary task performance, and increases the time users do not look on the road as well as the driver activity load.

APPARATUS

To evaluate our hypotheses, we developed an apparatus in the form of an instrument cluster (IC).

Hardware

To be able to compare the effect of different modalities, we first created an apparatus compatible with different display technologies. To test a glass-free technology, we used the lenticular autostereoscopic display built in a Toshiba Notebook P855-107. The 15.6" display has a resolution of 1366 x 768 in 3D mode and 1920 x 1080 in 2D mode. The display is equipped with a tracking unit that adjusts the perfect viewing angle based on the viewer's position. For a glass-based technology, we used the shutter display of an ASUS notebook G75VW equipped with Nvidia 3D Vision. This display has a screen size of 17.3" and a resolution of 1920 x 1080.

To track the user's gaze path for our study we use the Ergoneers Dikablis[2] glasses-based eye-tracker. We use Unity[3] with C# as scripting language to build our interactive UI.

User Interface

To reflect the potential of future UIs, we do not only employ standard UI elements commonly used in current instrument clusters for the purpose of our study, but integrate elements that take advantage of 3D space. This includes elements exploiting the spatial relationship between objects (e.g., a

[2]Ergoneers Dikablis: `http://www.ergoneers.com/en/products/dlab-dikablis/overview.html`

[3]Unity 3D - `www.unity.com`

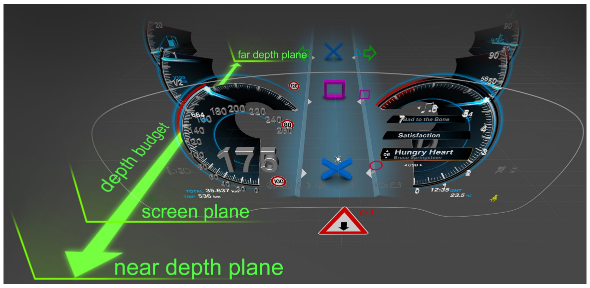

Figure 2: Top view of the implemented concept depicting the spatial arrangement of the displayed IC elements.

navigation system showing upcoming manoeuvres) and elements representing unexpected events (e.g., a warning that requires immediate action). Furthermore, we aim to design the IC in a way that allows elements to be structured and positioned in 3D space to reflect their current importance. For the depth layout we adhere to previously presented design guidelines [4]. The following UI elements are used (Figure 3):

Abstract Driving Space: The abstract representation of a street shows upcoming events such as navigation cues, traffic signs (speed limits), and traffic information (traffic jam). The depth position of the events inside this virtual space correlates with the actual distance from the vehicle. The Abstract Driving Space occupies the depth range from the screen plane to the maximum positive parallax (40 pixels).

Warnings: We integrate pop-up instructions showing urgent information the driver has to take immediately into account (e.g., collision warnings). The warning appears in front of the screen plane with a negative parallax of 20 pixels.

Further Elements: We also added common parts of an IC, including gauges (speed, RPM, temperature, fuel), indicator lamps, board computer, and a small infotainment menu. These elements are grouped on different layers (Figure 2).

To investigate the influence of UI complexity, we implemented two IC layouts (Figure 3). The first version depicts all UI elements described above. The second one shows a reduced information space with the most important elements (driving space, warnings, speed gauge, and rpm display).

DRIVING SIMULATOR STUDY

Using the apparatus we ran a driving simulator study with 56 participants to evaluate the impact of S3D on cognitive load.

Tasks

During the study participants need to perform a primary driving task (i.e., controlling the car) and a secondary task (i.e., reacting to expected and unexpected events). As primary driving task we use the following headway scenario. Participants need to follow a white vehicle on the right lane of a motorway. They are told to keep the same speed as the car in front (100 km/h) and maintain a constant distance of 50 m.

The secondary task is to estimate the depth relation of objects in the Abstract Driving Space. Symbols with varying shapes (circle, square, and triangle) appear at the end of the street and move towards the driver. For each symbol type a static

Figure 3: High (top) and low (bottom) complexity layout of the IC.

Figure 4: A button ‘X’ on the left side of the steering wheel needs to be pressed for judging the symbol’s position. A toggle button (‘↑’ or ‘↓’) on the right side of the steering wheel decodes the reaction on unexpected events.

“target zone” is marked with white arrows on the street (Figure 3) – the symbols on the right indicate the corresponding symbol type. When the symbol reaches its target zone (same depth position), the participant has to push a button ‘X’ on the steering wheel (Figure 4). If the button is pressed, the symbol disappears. When a new symbol appears, the system provides an auditory cue to make the driver aware of the new task. To make the task more difficult, the IC shows cross symbols as distractors. We chose this task because it requires frequent glances at the IC to check the current symbol positions.

Participants also need to react to unexpected events. Therefore, the IC shows a large icon with an arrow pointing upwards or downwards. According to the direction, participants should push the corresponding side of the toggle button on the right side of the steering wheel (Figure 4). We instruct the participants to react as accurate and fast as possible on these symbols. These events only appear if the depth judgment task requires an interaction. This constraint makes sure that the participant’s eyes are on the display when the warning appears. Thus, effects of focus switches can be eliminated.

Study Design

We use a mixed study design. Therefore, we apply one between and two within independent variables:

Visualization: We present participants a monoscopic (*2D*) and a stereoscopic (*3D*) visualization as within factor.

Complexity: To evaluate the influence of information complexity on the visualization we distinguish two variants. One shows all IC elements, providing a *high* visual complexity. The other one solely shows the IC elements that are necessary to solve the primary and secondary tasks (*low*). The participants experience both levels of information complexity during the study (within factor).

Technology: We use two display technologies in the study as between factor. One half of the participants solves the tasks on a shutter display (*shutter*), which requires 3D glasses but provides high 3D quality. To assess comparable results, the participants using the shutter display, wear the active shutter glasses during all conditions, even the monoscopic ones. The other half of the participants uses an autostereoscopic display (*autostereoscopic*) that lacks in 3D quality in terms of resolution and 3D artifacts like crosstalk.

Each within factor has two levels (four conditions). We counterbalance the order (Visualization, Complexity) using a latin square. Per condition we show each symbol ten times. In addition, we add two symbol tasks for each of the three symbol types that trigger a warning, one with an arrow pointing up and one with an arrow pointing down. One condition contains $3 \times (10 + 2) = 36$ tasks. The task order is randomized.

Setup

We evaluate both display technologies in the same driving simulator. The setup involves a basic simulator mock-up equipped with an adjustable driver seat, a multifunction steering wheel, as well as accelerator and brake pedals. Since the displays are both notebook screens, both are installed in a 45° angle behind the steering wheel, facing downwards. A surface-coated mirror is mounted on the notebooks' keyboard to reflect the displayed content to a vertical layer behind the steering wheel (Figure 1). Hence, the participants can see the displayed image at a position common for an IC.

The driving scene is displayed on a 52" LCD monitor with Full HD resolution (1920×1080 pixels) at 60 Hz. During the study, participants sit 2.5 meters in front of the driving scene and 75 cm away from the screen plane of the reflected IC.

Procedure

Participants were recruited through mailing lists. As they arrived in the lab we introduced the study procedure and provided a brief explanation about S3D. Participants first completed a stereo vision test based on random dot stereograms to qualify for the study. After completing a demographic questionnaire, the participants started with the first drive. This drive served as baseline without a secondary task to get familiar with the driving simulator and task. As soon as the participants felt comfortable with the driving task and controls, we introduced the secondary tasks. After we instructed the participants on how to react to appearing symbols and warnings, they practiced the reaction on 30 depth judgment and 6 warning tasks. Subsequently, the first of four test conditions started. During the conditions, participants had to complete the primary driving task and the secondary tasks simultaneously. For each test drive 30 symbol and 6 warning tasks appeared with a preceding training block of 5 symbol and 4 warning tasks. In total, each participant completed $5 + 4 + 30 + 6 = 45$ secondary tasks for each test drive. The task sequence started at a specified point on the test track that allows the participants to accelerate up to the required 100 km/h and to find a constant distance of 50 meters to the car in front. Each test drive lasted about 12 minutes and was followed by completing additional questionnaires. Finally, participants ranked the four conditions due to their preference and had the possibility to comment on the 3D effect and the information complexity. The study took about 90 minutes.

Metrics

We use six objective and subjective metrics to assess user performance and workload.

Primary Driving Task Performance

Since the experiment is conducted in a simulator, we measure driving performance. We consider the *longitudinal control* by measuring the standard deviation of the distance gap, that is, the distance between the rear-most surface of the lead vehicle and the forward-most surface of the following vehicle as mentioned in the 2013 SAE J2944 draft [10]. We evaluate the *lateral control* by measuring the standard deviation of the lateral position from the vehicle center to the right lane border.

Secondary Task Performance

Participants perform a concurrent secondary task. We log the *distance* between the actual position of a symbol and its respective target zone at the time when the participant pushes the left steering wheel button. For warning symbols we assessed task completion times (TCT) and error rates (ER).

Gaze Data

For the autostereoscopic group we use a remote eye tracker to track gaze. For technical reasons this measurement is not possible for the shutter group. We calculate the mean glance duration on the IC (i.e., mean eyes-off-the-road time) and switches between IC and road.

Intuitive Interaction (INTUI)

To measure the intuitiveness of the interaction with the UI, we use selected measures of the INTUI questionnaire [28]. Particularly we ask after each condition about *Effortlessness (E)*, *Gut Feeling (GF)*, and *Magical Experience (ME)*.

Driver Task Load (DALI)

Participants fill in a Driver Activity Load Index (DALI) questionnaire [24] after each condition. Using the proposed weighting procedure, a total score is calculated per condition that combines the ratings of the different workload aspects.

Simulator Sickness (SSQ)

To investigate details about simulator sickness, we provide the Simulator Sickness Questionnaire (SSQ) [16] to the participants. It measures the user's condition concerning nausea, oculomotor, disorientation, and a *total (sickness) score*.

		Autostereoscopic				Shutter			
		Presentation		Information		Presentation		Information	
Metric	Measure	2D	3D	Low	High	2D	3D	Low	High
Primary Task	**Lateral Control**	0.049 (0.018)	0.049 (0.019)	0.048 (0.014)	0.050 (0.022)	0.058 (0.016)	0.060 (0.016)	0.058 (0.016)	0.060 (0.016)
	Longitudinal Control	0.657 (0.366)	0.600 (0.311)	0.587 (0.302)	0.688 (0.375)	0.692 (0.352)	0.643 (0.335)	0.641 (0.319)	0.694 (0.366)
Secondary Task	**Mean Distance**	0.932 (0.323)	0.891 (0.302)	0.916 (0.291)	0.926 (0.359)	0.921 (0.324)	0.703 (0.262)	0.796 (0.300)	0.828 (0.327)
	SD Distance	0.746 (0.435)	0.608 (0.244)	0.636 (0.225)	0.726 (0.451)	0.615 (0.242)	0.544 (0.239)	0.568 (0.264)	0.592 (0.221)
	TCT	1.159 (0.177)	1.132 (0.147)	1.141 (0.136)	1.152 (0.186)	1.186 (0.190)	1.134 (0.142)	1.159 (0.192)	1.161 (0.145)
	Error Rate	0.009 (0.038)	0.025 (0.084)	0.013 (0.049)	0.021 (0.078)	0.018 (0.061)	0.018 (0.052)	0.012 (0.054)	0.024 (0.059)
Gaze Data	**Mean Glance**	0.860 (0.236)	0.907 (0.278)	0.853 (0.264)	0.914 (0.250)	-	-	-	-
INTUI	**Effortlessness**	4.739 (0.801)	4.950 (0.673)	4.983 (0.784)	4.764 (0.752)	4.689 (1.131)	4.793 (1.024)	4.975 (1.093)	4.507 (1.013)
	Gut Feeling	3.616 (1.048)	3.705 (1.169)	3.599 (1.157)	3.643 (1.127)	3.777 (1.200)	3.589 (1.204)	3.598 (1.203)	3.768 (1.202)
	Magical Experience	3.920 (1.027)	4.509 (1.168)	4.177 (1.138)	4.321 (1.183)	3.835 (1.108)	4.893 (0.957)	4.205 (1.223)	4.522 (1.080)
DALI	**Total Score**	3.023 (0.858)	3.151 (0.939)	3.025 (0.901)	3.099 (0.925)	2.906 (1.078)	3.077 (1.028)	2.941 (1.043)	3.041 (1.068)
SSQ	**Total Score**	0.172 (0.177)	0.160 (0.158)	0.167 (0.189)	0.158 (0.142)	0.140 (0.137)	0.153 (0.177)	0.140 (0.142)	0.153 (0.173)
Ranks	**Total Score**	2.732 (1.036)	2.268 (1.168)	2.448 (1.111)	2.589 (1.141)	3.161 (0.848)	1.804 (0.903)	2.339 (1.049)	2.625 (1.153)

Table 1: Mean and SD values (in brackets) for all measures and conditions. Note that the measuring of gaze data was not possible for the shutter technology.

RESULTS

56 participants (11 female, 45 male) aged between 20 and 59 ($M = 32.75$, $SD = 8.96$) took part in this study. All had normal or corrected to normal visual acuity and passed the stereo vision test. For analyzing the task performance measurements as well as the questionnaires, we use two-way mixed Analysis of Variance (ANOVA) since Kolmogorov-Smirnov tests show that the data is normally distributed. Pertaining the subjective ranking we analyze the data with a Friedman analysis of variance. Table 1 shows all means and standard deviations.

Primary Driving Task Performance

Testing 2D against 3D shows no statistically significant differences regarding the primary driving task, in terms of lateral, $F(1, 54) = 2.297$, $p = .135$, and longitudinal control, $F(1, 54) = 2.869$, $p = .096$. The two display technologies (*autostereoscopic* and *shutter*) show a statistically significant influence on the primary task, in terms of lateral control, $F(1, 54) = 6.102$, $p = .017$, $r = .393$, to the advantage of the autostereoscopic presentation. Regarding the standard deviation of the distance to the preceding car, we cannot show a statistically significant difference for the evaluated technologies, $F(1, 54) = .296$, $p = .588$. The IC information complexity has a statistically significant effect on the primary driving task concerning the longitudinal control, $F(1, 54) = 5.917$, $p = .018$, $r = .386$. However, we cannot show any statistically significant difference for the lateral control, $F(1, 54) = 2.353$, $p = .131$, pertaining complexity.

Secondary Task Performance

The 3D instrument cluster version shows advantages over its 2D counterpart for the secondary tasks. For the depth judgment task, we analyze means and standard deviations for the distance between symbol and target zone. Users judge the positions of the symbols more accurately for 3D regarding mean, $F(1, 54) = 21.503$, $p < .001$, $r = .650$ and standard deviation, $F(1, 54) = 11.740$, $p < .001$, $r = .527$. The two levels of Complexity and Technology do not reveal statistically significant effects for the mean and standard deviation of the judgments. However, the interaction effect for Visualization * Technology regarding mean distance between symbol and target zone is statistically significant, $F(1, 54) = 10.031$, $p < .003$, $r = .494$. S3D does not improve depth judgment for autostereoscopy but for shutter (Figure 5).

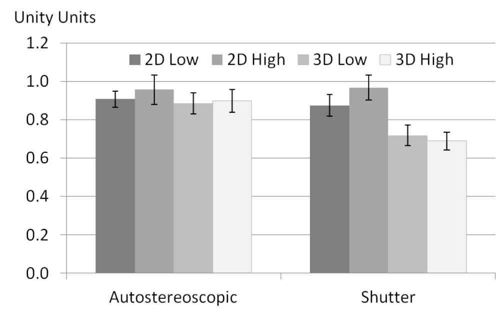

Figure 5: The mean and standard errors for the mean distance between symbol and target zone show that user performance improves when solving the task in S3D on the shutter display compared to autostereoscopy.

Comparing 2D against 3D, there is a significant difference for TCT when reacting on unexpected instructions, $F(1, 54) = 7.726$, $p = .007$, $r = .372$, to the advantage of 3D. Neither the used display technologies, $F(1, 54) = .160$, $p = .691$, nor the degrees of information complexity, $F(1, 54) = .204$, $p = .653$, show significant effects for the measurement of TCT. In general, the participants reacted very accurately on the unexpected instructions. Since the data of the error rates are not normally distributed, we used a Friedman ANOVA. These tests show no statistically significant differences for the autostereoscopic, $X^2(3) = 1.320$, $p = .724$, as well as for the shutter technology, $X^2(3) = 2.455$, $p = .484$. The error rate is highest for the autostereoscopic display with the 3D visualization of the high information content ($ER = 3\%$) and less in the condition $3D_{low}$ ($ER = 2\%$), $2D_{high}$ ($ER = 1\%$), and $2D_{low}$ ($ER = 1\%$). The shutter sample has the highest ER for the conditions $2D_{high}$ ($ER = 2\%$) and $3D_{high}$ ($ER = 2\%$) whereas $2D_{low}$ ($ER = 1\%$) and $3D_{low}$ ($ER = 1\%$) has the lowest ER.

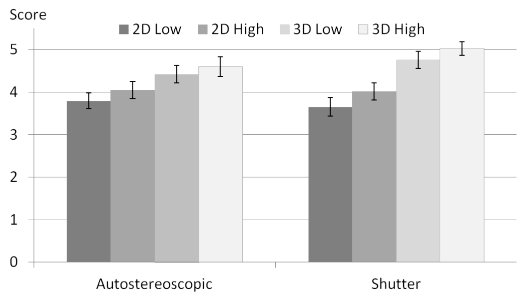

Figure 6: The mean and standard errors for the dimension *Magical Experience* show that the shutter technology provides greater scores for 3D visualizations than the autostereoscopic display.

Gaze Behavior

For gaze behavior we measure the mean glance duration onto the IC for the autostereoscopic sample. The visualization levels do not show a significant influence on gaze behavior concerning the mean glance duration, $F(1, 27) = 3.691$, $p = .065$. However, there is a statistically significant effect of the different levels of information complexity for the mean glance duration on the IC, $F(1, 27) = 9.645$, $p = .004$ $r = .486$, to the benefit of the low information level.

Intuitive Interaction

Effortlessness is statistically significant for information complexity, to the benefit of the low complexity level, $F(1, 54) = 15.296$, $p < .001$, $r = .581$, but not for Visualization, $F(1, 54) = 3.536$, $p = .065$, and the Technology, $F(1, 54) = .251$, $p = .618$. *Gut Feeling* does not reveal any significances for the tested variables. In contrast, *Magical Experience* shows significant effects for Visualization, $F(1, 54) = 65.236$, $p < .001$ $r = .835$, Complexity, $F(1, 54) = 5.618$, $p = .021$, $r = .364$, and the interaction of Visualization * Technology, $F(1, 54) = 5.282$, $p = .025$, $r = .376$. Figure 6 illustrates that 3D is superior to 2D and that the high information density gains advantage over the low information degree for this dimension. Moreover, the 3D effect of the shutter technology provides larger scores than autostereoscopy.

Driver Task Load

The DALI reveals statistically significant differences for Visualization, $F(1, 54) = 7.795$, $p = .007$, $r = .967$, to the benefit of 2D. The results show no statistically significant differences for Technology and information load.

Simulator Sickness

The results of the SSQ are not statistically significant for the Visualization, $F(1, 54) = .002$, $p = .967$, Technology, $F(1, 54) = .251$, $p = .618$, and Complexity, $F(1, 54) = .002$, $p = .956$.

Rankings

A Friedman test for the rankings of the different conditions is not statistically significant for the autostereoscopic sample, $X^2(3) = 4.929$, $p = .177$, but for the shutter sample, $X^2(3) = 33.347$, $p < .001$. Pairwise Wilcoxon tests with Bonferroni corrections show that the participants rate all 3D conditions significantly better than 2D, $p < .006$.

DISCUSSION

The study reveals insights into the influence of visualization, technology, and complexity when introducing 3D displays to automotive UIs. In the following, we discuss the findings.

H1: Influence of Visualization

The findings from our study yield – at a first glance – no influence of an S3D visualization on the primary driving task. We assume that two contradicting effects may have superimposed each other and, hence, obscure a potential main effect. On one hand, our log data shows that S3D increases the user's secondary task performance, which, in turn, allows more attention to be directed towards the primary driving task. Participants feel supported in filtering the relevant information due to the 3D effect. They state that the spatial arrangement of the items "*clarifies priorities*", "*fosters the display's structure*", and provides an "*intuitive understanding of spatial relations*". On the other hand, 3D artifacts and the accommodation-convergence mismatches can induce eye strain, dizziness, or headache that can negatively affect the primary driving task. But the results of the SSQ do not show a significant effect on the drivers' condition. However, participants subjectively rate S3D as more distracting than 2D. The participants' comments justify the perceived distraction through the "*fascinating*" character of S3D and induced visual load. However, our gaze measures do not show a significant increase of the visual load.

With regard to the secondary driving task, the 3D visualization supports depth judgments and decreases TCT for warnings. The shutter sample reveals an average decrease of 53 ms for highlighting the warning with binocular disparity. This means, the breaking distance can be reduced by 1.4 m. Former research already shows that binocular highlighting, decreases search times [13].

Furthermore, S3D has a positive influence on attractiveness, UX, and acceptance. Participants state the 3D effect to be "*stylish*", "*attractive*", "*fascinating and innovative*". This conforms to former comparisons of 2D/3D presentations for gaming [26], automotive [5], and mobile applications [27].

H2: Influence of Technology

As our results show, the use of autostereoscopy has a positive impact on the primary driving task. Yet, participants state that the glasses are "*annoying*". The decrease in driving performance for shutter is attributable to the flickering and the decrease in brightness while wearing the active glasses.

The shutter technology increases secondary task performance in terms of the accuracy in making depth judgments. This result corresponds to the findings of Alpaslan et al. [2] showing better task performances for shutter compared to autostereoscopic displays. We assume that the reduced quality of the autostereoscopic display decreases the accuracy of judging depth as applied in our study. Since the used autostereoscopic display shows artifacts like crosstalk, particularly tasks that require the perception of higher parallaxes are affected as it is the case for the depth judgment task. A second reason is the reduced resolution of the autostereoscopic technology, that probably affects the accuracy in judging depth.

	Visualization (2D vs. 3D)	Technology (Autostereo vs. Shutter)	Complexity (Low vs. High)
Primary Task Performance	n.s.	Lateral control is better for autostereoscopic displays.	Longitudinal control is better for UIs with lower complexity.
Secondary Task Performance	3D lowers TCT for unexpected events, and allows more accurate judging of positions.	n.s.	n.s.
Gaze Performance	n.s.	–	The mean glance duration is lower for UIs with low complexity.
INTUI	Magical Experience is better for 3D.	A shutter display maximizes the Magical Experience for 3D visualizations.	Low complexity UIs have less Magical Experience, but better scores for Effortlessness than higher complexity UIs.
Driver Activity Load	2D visualizations subjectively lower distraction.	n.s.	n.s.
Simulator Sickness	n.s.	n.s.	n.s.
Acceptance	3D visualizations are preferred over 2D presentations.	n.s.	n.s.

Table 2: Overview of the results of our experiment. Note that measuring gaze data was not possible for the shutter technology.

Finally, the display technology has a significant influence on UX and user acceptance. Again, the superior 3D quality of the shutter technology due to the higher display resolution and the absence of crosstalk leads to better subjective ratings for the attractiveness and the preference of the 3D effect.

H3: Influence of Complexity

The degree of information complexity has a significant effect on the primary driving task in terms of longitudinal control. Moreover, a reduced visual complexity decreases eyes off the road time and results in better subjective ratings of effortlessness. However, the participants mentioned the 3D effect to "*declutter the display*" and "*improve clarity for high information densities*" compared to 2D.

Although our study shows that higher information complexities reduce performance and increase driver distraction, participants perceive the high information density as more attractive. We assume that higher information densities foster the drivers' confidence and provides a feeling of control.

LIMITATIONS

We acknowledge the following limitations. First, the perceived brightness of the display is not consistent across conditions because of the darkening factor of the shutter glasses. This could have an influence on user performance. Second, we conducted our study in a driving simulator. While this increases internal validity (e.g., we are able to control the traffic), it reduces the external validity. However, we deliberately opted for this setting, as we needed a highly controllable environment while not putting participants at risk by having them drive on a real motorway. Third, the proposed (secondary) tasks are artificial. Even though these tasks would not be performed in real world driving scenarios they are quite similar to tasks like responding to routing instructions, navigation cues or urgent alerts as warnings and notifications. We believe that the chosen tasks allow for transferring the results to various use cases of automotive UIs.

CONCLUSION

In this paper, we investigate the influence of S3D on the user's primary and secondary task performance as well as cognitive load in an automotive setting. We developed an interactive 3D instrument cluster with two levels of complexity and conducted a driving simulator study to evaluate the impact of a stereoscopic visualization, the used 3D display technology, and the level of complexity on the driver. The core findings of our study provide designers of future S3D UIs specific hints how to optimally support the driver:

Use S3D to Increase Secondary Task Performance: Users perform better in secondary tasks using 3D visualizations. Since the drivers judge spatial relations between UI objects more accurately, S3D can enhance UI elements that, e.g., represent the distance to a preceding car or navigation cues. Also, highlighting instructions using S3D shortens interaction times. Thus, using S3D is advisable to highlight urgent content, (e.g., warnings and notifications).

Choose an Appropriate Display Technology: Autostereo displays are more suited for automotive UIs, since glasses are disturbing and reduce the primary task performance. However, the quality of the 3D effect is crucial for secondary task performance. Thus, we suggest that autostereoscopic displays should exhibit a 3D quality comparable to state-of-the-art shutter technology for a successful integration into cars.

Consider the Complexity of the Displayed Content: Higher information complexities reduce primary task performance and increase distraction. At the same time, they make UIs more attractive. Structuring information on different depth layers using S3D reduces the perceived complexity by decluttering the content.

Despite the limitations of our study, the outcomes reveal that S3D interfaces offer benefits compared to their 2D counterparts. We believe that the chosen tasks let us transfer the results to various automotive use cases (e.g., ACC, navigation, warnings, notifications, etc.). As a next step we plan to implement real automotive use cases to verify this translation.

Finally, designing S3D UIs is in many ways challenging particularly with regard to the security-critical automotive context that requires the user to engage in parallel tasks. The findings of this research support the development of reasonable S3D UIs and points towards aspects that influence their successful application in cars.

ACKNOWLEDGEMENTS
We thank Robert Wilde for his support in this project.

REFERENCES
1. Alliance of Automobile Manufacturers. Statement of principles, criteria and verification procedures on driver interactions with advanced in-vehicle information and communication systems, 2006.
2. Alpaslan, Z., Yeh, S., Rizzo, A., and Sawchuk, A. Quantitative comparison of interaction with shutter glasses and autostereoscopic displays. In *Proc. SPIE* (2005).
3. Bach, K., Jæger, M., Skov, M., and Thomassen, N. Interacting with in-vehicle systems: Understanding, measuring, and evaluating attention. In *Proc. HCI '09*, BCS (2009).
4. Broy, N., Alt, F., Schneegass, S., Henze, N., and Schmidt, A. Perceiving layered information on 3D displays using binocular disparity. In *Proc. PerDis '13*, ACM (2013).
5. Broy, N., André, E., and Schmidt, A. Is stereoscopic 3D a better choice for information representation in the car? In *Proc. AutomotiveUI '12*, ACM (2012).
6. Broy, V., Althoff, F., and Klinker, G. iflip: a metaphor for in-vehicle info systems. In *Proc. AVI'06*, ACM (2006).
7. Brunken, R., Plass, J. L., and Leutner, D. Direct measure- ment of cognitive load in multimedia learning. *Edu. Psych. 38*, 1 (2003).
8. Collet, C., Delhomme, G., and Dittmar, A. Subjective aspects of mental workload in air-traffic control, 2003.
9. Gopher, D., and Braune, R. On the psychophysics of workload. *Human Factors 26*, 5 (1984).
10. Green, P. Standard definitions for driving measures and statistics: Overview and status of recommended practice J2944. In *Proc. AutomotiveUI '13*, ACM (2013).
11. Häkkilä, J., Posti, M., Koskenranta, O., and Ventä-Olkkonen, L. Design and evaluation of mobile phonebook apps with stereoscopic 3D user interface. In *CHI '13 EA*, ACM (2013).
12. Hart, S. G., and Stavenland, L. E. Development of NASA-TLX: Results of empirical and theoretical research. In *Hum. Ment. Workl.* Elsevier, 1988.
13. Huhtala, J., Karukka, M., Salmimaa, M., and Häkkilä, J. Evaluating depth illusion as method of adding emphasis in s3d mobile displays. In *Proc. MobileHCI '11*, ACM (2011).
14. International Organization for Standardization. ISO 16673: Road vehicle – ergonomic aspects of transport information and control systems-occlusion method to assess visual demand due to the use of in-vehicle-systems, 2007.
15. Jahn, G., Oehme, A., Krems, J. F., and Gelau, C. Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study. *Tran. Res. Part F 8*, 3 (2005).
16. Kennedy, R. S., Lane, N. E., Berbaum, K. S., and Lilienthal, M. G. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *IJAP 3*, 3 (1993).
17. Krüger, K. *Nutzen und Grenzen von 3D-Anzeigen in Fahrzeugen*. PhD thesis, HU Berlin, 2007.
18. Lambooij, M., Fortuin, M., Heynderickx, I., and IJsselsteijn, W. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science* (2009).
19. Mattes, S. The lane-change-task as a tool for driver distraction evaluation. *Most* (2003).
20. McIntire, J. P., Havig, P. R., and Geiselman, E. E. What is 3D good for? a review of human performance on stereoscopic 3D displays. vol. 8383 (2012).
21. Mizobuchi, S., Terasaki, S., Häkkinen, J., Heinonen, E., Bergquist, J., and Chignell, M. The effect of stereoscopic viewing in a word-search task with a layered background. *Journal of the Society for Information Display 16*, 11 (2008).
22. Mousavi, S. Y., Low, R., and Sweller, J. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Edu. Psy. 87*, 2 (1995).
23. Paas, F., Tuovinen, J. E., Tabbers, H., and Van Gerven, P. W. Cognitive load measurement as a means to advance cognitive load theory. *Edu. Psych. 38*, 1 (2003).
24. Pauzie, A. A method to assess the driver mental workload: The driving activity load index (DALI). *Humanist 2* (2008).
25. Roscoe, A. H. Heart rate as a psychophysiological measure for in-flight workload assessment. *Ergonomics 36*, 9 (1993).
26. Schild, J., LaViola, J., and Masuch, M. Understanding ux in stereoscopic 3D games. In *Proc. CHI '12*, ACM (2012).
27. Sunnari, M., Arhippainen, L., Pakanen, M., and Hickey, S. Studying ux of autostereoscopic 3D menu on touch screen mobile device. In *Proc. OzCHI '12*, ACM (2012).
28. Ullrich, D., and Diefenbach, S. Intui. exploring the facets of intuitive interaction. In *Proc. M&C* (2010).
29. Urey, H., Chellappan, K., Erden, E., and Surman, P. State of the art in stereoscopic and autostereoscopic displays. *Proc. IEEE 99*, 4 (2011).
30. Van Beurden, M. H., Ijsselsteijn, W. A., and de Kort, Y. A. Evaluating stereoscopic displays. vol. 7863 (2011).
31. Van Orden, K. F., Limbert, W., Makeig, S., and Jung, T.-P. Eye activity correlates of workload during a visuospatial memory task. *Human Factors 43*, 1 (2001).
32. Wittmann, D., Baier, A., Neujahr, H., Petermeier, B., Sandl, P., Vernaleken, C., and Vogelmeier, L. Development and evaluation of stereoscopic situation displays for air traffic control. *Proc. Innovation in Mechanical Engineering* (2011).