Gaze-based interaction on multiple displays in an automotive environment

Tony Poitschke, Florian Laquai, Stilyan Stamboliev, and Gerhard Rigoll
Institute for Human-Machine-Communication
Technische Universität München
Theresienstr. 90, 80333 Munich, Germany
{poi,laq,sta,ri}@mmk.ei.tum.de

Abstract—This paper presents a multimodal interaction system for automotive environments that uses the driver's eyes as main input device. Therefore, an unobtrusive and contactless sensor analyzes the driver's eye gaze, which enables the development of gaze driven interaction concepts for operating driver assistance and infotainment systems. The following sections present the developed interaction concepts, the used gaze tracking system, and the test setup consisting of multiple monitors and a large touchscreen as central interaction screen. Finally the comparison results of the gaze-based interaction with a more conventional touch interaction are being discussed. Therefore, well-defined tasks were completed by participants and task completion times, distraction and cognitive load were recorded and analyzed. The tests show promising results for gaze driven interaction.

Index Terms—Multimodal interaction, gaze tracking, touch interaction, multi display interaction, automotive

I. INTRODUCTION

The rapid advances in the field of in-car information systems (both driver assistance and driver information) contribute to an increasing mental workload of drivers originating from secondary tasks. This actual condition will even worsen within the coming years. To defuse this situation, vehicle manufacturers as well as research institutions are increasingly developing new intuitive and easy to use interaction concepts that lower the driver's distraction from the driving task while operating. In order to achieve a less distracting operation of in-vehicle systems, this contribution evaluates the potential of using an eye tracker to simplify human-machine interaction in automotive environments.

Today, eye tracking systems are widely used in many scientific and application-oriented disciplines. These include areas of medicine, neurology, usability, and also cognitive sciences and psychology. Concerning the field of human-machine interaction, such systems are used as a novel interaction modality that enables for example gaze controlled applications in the aid of physically challenged patients. Thus, several handicapped authors managed to write a book using gaze-operated on-screen keyboards [7]. Furthermore, recent developments in gaze tracking technology (e.g., reduced installation space, high reliability, data rates, and accuracy, etc.) allow for an integration of eye tracking system into safety critical surroundings like trains, airplanes, and automobiles.

The presented prototype bases on the eye tracking system presented in [11], [12], and the gaze data handling system presented in [1]. These systems provide various eye parameters
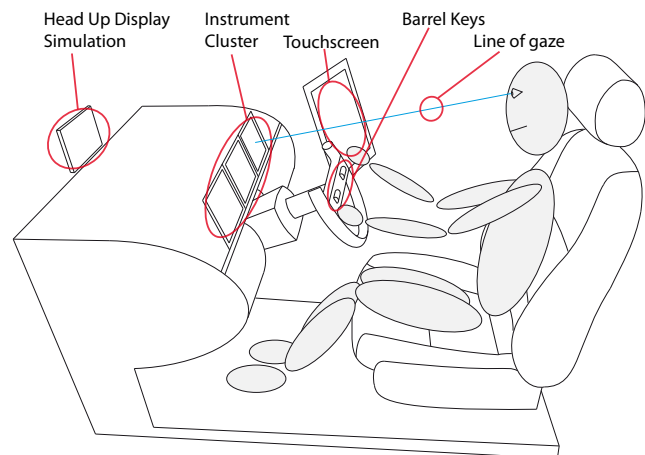

Fig. 1. Mockup for user evaluations depicting display arrangement and interaction devices

like eye position, gaze vector, gaze point, fixation data, etc. The derived data are applied to implement gaze-driven interaction strategies for simple system control tasks, for example a gaze-based selection of menu items that are manipulated via analog input devices, e.g. a barrel key on the steering wheel.

II. GAZE INTERACTION CONCEPT

The large potential to reduce required interaction time is based on the fact that gaze direction (visual attention) is highly correlated with the user's mental attention, interests, plans, and intentions [6], [15]. As scanning the environment is part of the natural orientation and task planning process, eye movements and especially fixations (e.g., the fixation of a menu item on the screen) can be used for selecting visible objects in the user's surroundings. Thus, the selection process can be accomplished very quickly and easily to understand even for untrained users by simply looking at the desired menu item or button. Hence, the driver can keep his hands on the steering wheel which is fundamental for optimal driving safety.

However, in the context of gaze-based interaction the "Midas Touch" problem described in [5] is a big challenge. An additional input channel for confirming an action (e.g., speech input, hardware button, etc.) is an appropriate method to cover this problem which occurs due to the fact that evolution
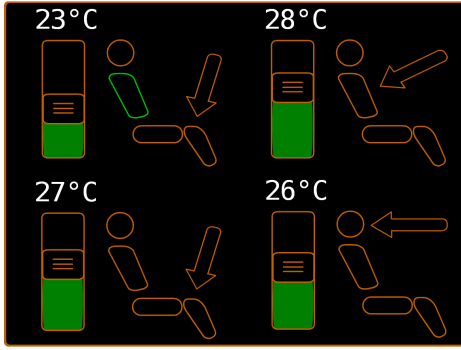
Fig. 2.　Climate control widget for gaze and touch interaction

Fig. 3.　Moving widgets to other displays with a touch gesture, driver view

optimized the human visual system as a sensor, not an actor. Furthermore, it is a very important design issue especially in an automotive environment to consider that an input action has to be interruptible. It can happen at any time that the driver must interrupt an input procedure for an urgent traffic situation. Consequently this means that gaze tracking cannot be used to confirm a previous input e.g. by using conventional dwell time techniques.

Based on these findings, gaze-driven input is well suited for the task of object selection. Thus, the proposed concepts combine gaze-based object selection with a hardware interface (multi-functional barrel keys on the steering wheel) for confirmation and value adjustments. As described in [8] a haptic interface accounts for lowest distraction from the main driving task compared to other interaction styles like touch screens. These keys serve as a manipulation input device by rotating them up and down which can be used for list scrolling or value adjustment. This can be the desired distance for an ACC (Adaptive Cruise Control) system or the temperature of the climate control system. Also they are used for confirmation actions by pressing them, e.g. selection of a sub menu and entering it. Hence, combining eye gaze and barrel keys results in a mouse-like interaction style: pointing at a desired item (gaze) and activating the appropriate functionality (e.g., pressing the barrel key for a selection or rotating it to scroll a list). The climate control widget shown in figure 2 can be for example controlled by looking at a slider and turning the barrel keys to adjust the respective temperature setting.

Additionally the selected object is highlighted by changing its color to achieve an intuitive interaction and to avoid the impact of the "Midas Touch" problem. Therefore, the proposed system implements a color change from the standard orange color to green in order to visualize that a menu item is preselected but not activated (see fig. 2 top left). This visual feedback occurs exactly at the position the user is looking at and is therefore very straightforward.

The used gaze tracker provides an accuracy of $0.45\,°$ for the reconstructed gaze direction [11], [12]. This results in an average spatial error of ca. 6mm on the displays. The driver's current point-of-gaze in display coordinates is calculated from the eye positions and gaze vectors according to [1].
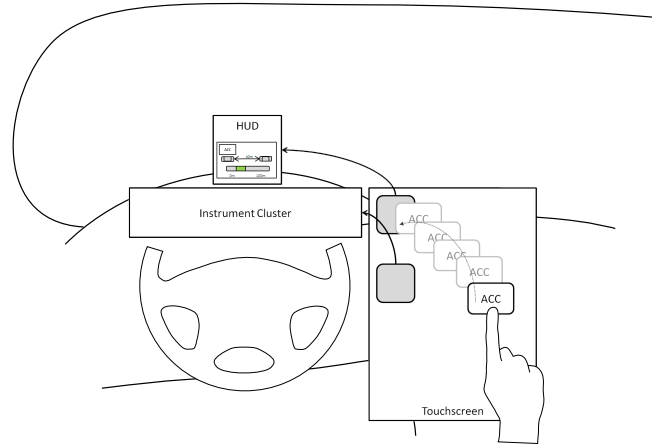
III. MULTIPLE DISPLAY CONTROL

The above described selection method can be extended from one central interaction screen to several displays that are distributed over the whole instrument cluster area. Instruments that are currently still implemented as physical gauges or switchable icons can be completely replaced by large digital screens. These are combined to create a practically seamless display area of which a portion is touchable. But the screens can be used for more than a mere replacement for conventional instruments, it is desirable to use this area in a more dynamic way.

Common output and control elements can be interpreted like widgets of a graphical interface and in consequence enable a nearly arbitrary arrangement. As a result the problem arises to move widgets to displays that are not touch-enabled and back again to the main display. This can be solved by reserving display space on the touchscreen which represents the other displays, see figure 3. That way smaller representations of the widgets (icons) can be moved from a source area to a remote display or the main workspace on the touchscreen as seen in the screenshot in figure 4.

Despite the statements in the previous paragraphs (see also [8]) about the high visual attention demand of touch interaction, the functionalities to move widgets to different displays was still implemented with touch input. This decision against gaze driven interaction was made as this task would require a relatively long interaction sequence to accomplish compared to a simple object selection. A reconfiguration would ideally be done while the car is standing still due to the distraction-prone nature of this task.

All widgets also support pure touch interaction besides gaze interaction. This provides the conditions for being able to compare these two interaction techniques with exactly the same functions and the same evaluation setup.
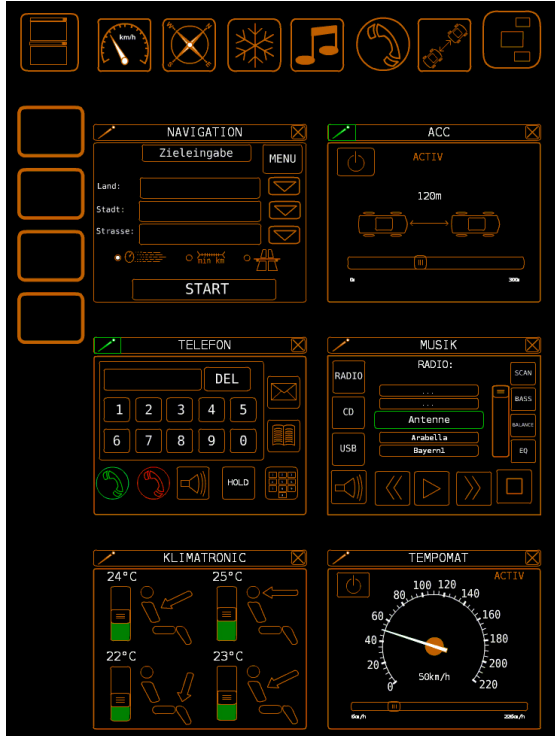
Fig. 4. Screenshot of the touchscreen GUI, 3 widgets on the workspace, widget source on the upper edge, remote screen representations on the upper left edge

Fig. 5. Driver view of the Lane Change Task with instruction signs and racing line, from [9]

IV. USER EVALUATION

As mentioned before the two interaction concepts shall be compared in a user study regarding Total Task Time and Cognitive Load while accomplishing a simple driving task. Additionally a usability rating is done by the participants.

A. Test setup

A mockup as seen in figure 1 was used with a 52" monitor positioned at 2m distance in front of the user's head. For the instrument cluster displays 3 flat screens with 8" size were placed directly behind the steering wheel and the Head-Up Display was installed 1.25m away from user's head by means of a 12" flat screen. The touchscreen is a 15" screen mounted in portrait orientation.

For the primary driving task the Lane Change Task (LCT, see [9]) was used. It is a widely used evaluation tool with simple driving scenarios primarily for evaluation of secondary tasks. The participants drive on a straight 3 lane road without foreign vehicles. Every 140 - 188m signs are installed which provide lane changing instructions for the driver, see figure 5. The maximum speed is limited to $60\frac{km}{h}$ and drivers were instructed to go with highest possible speed. This ensures that the primary driving task is simplified to a lateral control task. Also they were instructed to perform the lane change as soon as the signs are visible, which happened 60m before the signs. The L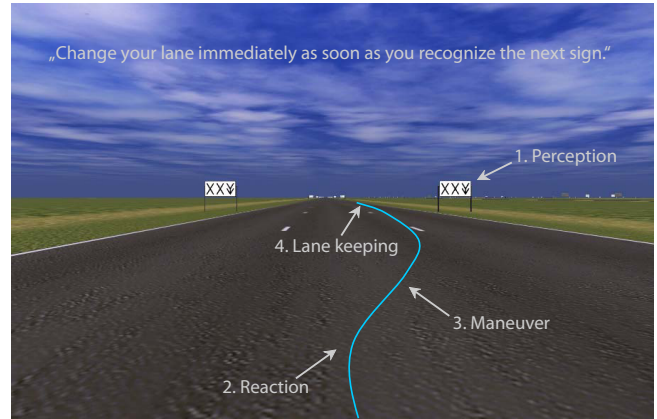CT creates a log file with timestamp, position, speed etc. which allows for later analysis of the deviation from the proposed trapezoidal racing line.

Further a Peripheral Detection Task (PDT, see [2]) must be accomplished to assess the visual and mental distraction caused by the secondary interaction tasks. In our evaluation only one point positioned $10°$ left of the neutral line of gaze was displayed every 3 - 6 seconds for a duration of 2 seconds. The detection of the point had to be acknowledged by pressing one of two possible buttons (for left-handed or right-handed drivers) on the steering wheel. All relevant points in time (point display, button press etc.) are also recorded for later analysis.

B. Test procedure

For the evaluation 15 participants (12 male, 3 female) with an average age of 23.5 years conducted the procedure. They all drive cars infrequently (87% drive less than 10 000 km per year) and 12 of them have used touchscreens before. The test procedure consisted of the following steps, all driving courses took ca. 8 minutes:

- Questionnaire about demographic properties (age, gender, previous experience etc.)
- Driving with PDT only, baseline
- Driving with PDT and touch interaction
- Driving with PDT and gaze interaction
- Driving with PDT only, for assessment of learning effects
- Questionnaire containing "AttrakDiff" for subjective impressions, Usability related questions and SEA scale

The AttrakDiff questionnaire is designed to obtain a rating for the hedonic and pragmatic quality of a system or product. It uses the semantic differential (word pairs with opposite meanings) with a scale between the words with 7 steps. Hedonic quality "Indicates to what extent the functions of a product enhance the possibilities of the user, stimulate him or communicate a particular identity (e.g. by creating a professional, cool, modern or any other impression). Hedonic quality is divided into two sub-qualities, namely identity (HQ-I) and stimulation (HQ-S). HQ-I indicates how well a user identifies with the product while HQ-S indicates the extent to

which a product supports the needs of a person to develop and move forward by offering novel, interesting and stimulating functions, contents, interactions and styles of presentation" (from [16]). Pragmatic quality (PQ) is an indicator for usability and therefore how well a user achieves his goals with the system. The fourth index is the attractiveness (ATT) of a system based on quality perception.

An estimation of the subjective mental workload, or how mentally demanding a system is, can be obtained with the SEA scale. It is a one-dimensional scale ranging from 0 (not demanding at all) to 110 (exceedingly demanding) on which participants mark a single point that matches their own estimation.

The tasks for both touch and gaze driven interaction are as follows:

- Set the distance of the ACC (Adaptive Cruise Control) to 110 meters
- Set target speed of cruise control to $100 \frac{km}{h}$
- Set climate control on the upper right to $30\,°C$ and air vents towards the feet
- Go to Navigation - Favorites - Theresienstrasse
- Set target speed of cruise control to $80 \frac{km}{h}$
- Set the distance of the ACC (Adaptive Cruise Control) to 200 meters
- Set climate control on the lower left to $20\,°C$ and air vents towards the upper part of the body
- Set target speed of cruise control to $80 \frac{km}{h}$
- Set the distance of the ACC (Adaptive Cruise Control) to 150 meters
- Go to Navigation - Target entry - Netherlands, Highways preferred - Start

All widgets were positioned automatically for the participants to avoid extensive distraction due to the widget distribution task as mentioned before. The arrangement was as follows: Cruise Control and tachometer - Head-Up Display, ACC - left display in instrument cluster, climate control - middle display, navigation - right display, the main workspace was located on the touchscreen.

V. EVALUATION RESULTS

In the following section the evaluation results shall be discussed. Two of the participants already had previous experience with gaze driven interaction and are (where appropriate) treated separately in the analysis.

Figure 6 shows the overall rating of both systems in a coordinate system with pragmatic and hedonic quality. The centers of the rectangles point to the average rating while the transparent borders depict the confidence intervals and thus the agreement of the test persons to the overall estimation. The smaller the confidence intervals are, the more all participants agreed to the average rating. As we can see the gaze interaction method is clearly more desired than the touch method.

Also figure 7 shows the superior results of the gaze driven interaction in all four dimensions. Especially for the HQ-S and attractiveness ratings (ATT) the gaze system has a clear advantage against touch interaction.
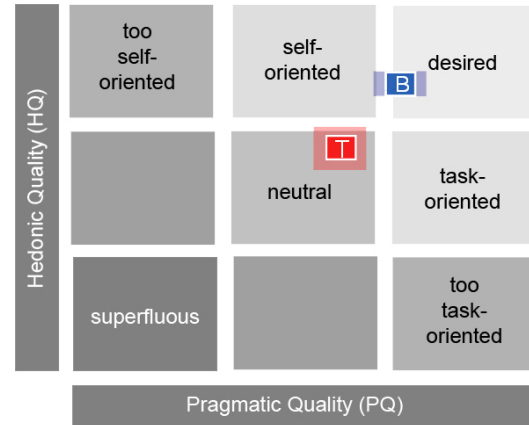

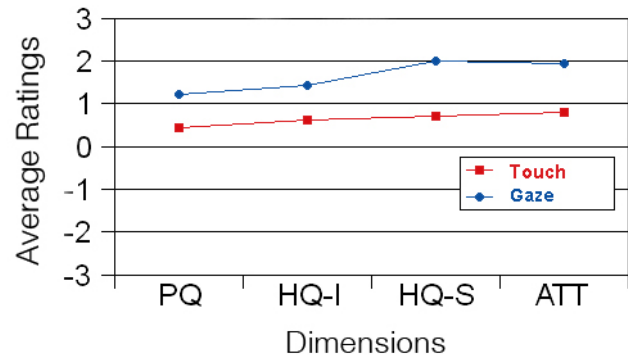Fig. 6.   AttrakDiff overvall results (T: Touch input, B: Gaze input)

Fig. 7.   AttrakDiff results in greater detail including the four main dimensions

The participants were also asked to answer the questions from figure 8, where 5 means total agreement and 1 means total disagreement. Due to the relatively small number of participants, a Box-Whisker-Plot was used. The green box shows the intervals that contain the rating of 50% of the participants who rated closest to the median which is indicated by the red line. Outliers are marked with the black whiskers. Participants rated the gaze driven interaction not distinctively better than a conventional input method (4th question). When asked about the visibility of the road during interaction, gaze driven input got clearly better results than touch driven input (question 4 and 5).

Also a rating of the subjectively perceived effort (SEA scale, see [4]) was produced as depicted in figure 9. In the box plot the interquartil ranges are denoted by the boxes, the median is the thick line insde the boxes and the whiskers denote the maximum and minimum ratings. It can be seen that the medians and boxes are clearly separated and thus show a less demanding rating for the gaze interaction.

Besides the questions about subjective ratings, the participants were also asked to conduct the tasks described in the previous section. For each task a difficulty level was calculated using the following formula:
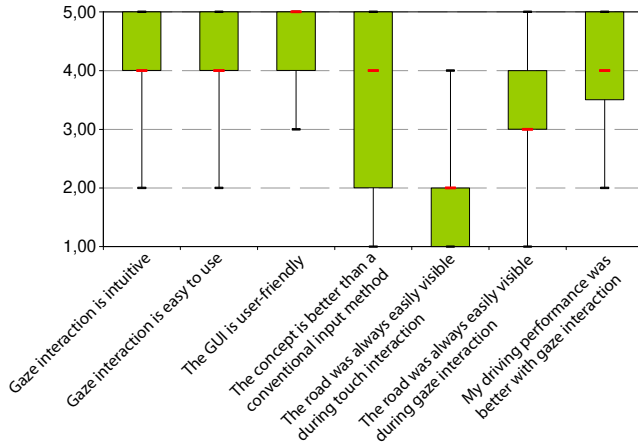
Fig. 8. Subjective rating of both interaction methods, 5=totally agree, 1=totally disagree
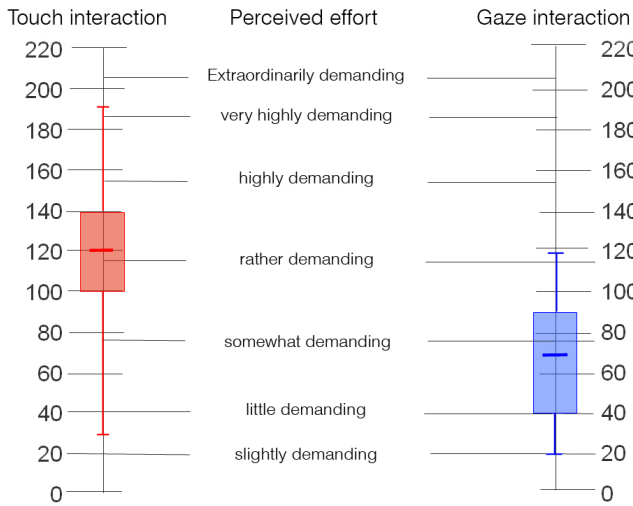

Fig. 9. Rating of subjectively perceived effort

$$Difficulty = Number\ of\ steps \cdot \frac{Window\ size}{Smallest\ object\ size}$$

For example the task "Go to Navigation - Favorites - Theresienstrasse" has 2 steps (selecting the Favorites button and Scrolling in a list to the entry Theresienstrasse), the window size is 1 because it spans over the whole display area and the smallest object, which is the Favorites button, is 1/4 of the display area. Hence the difficulty for this situation is "8" from $2 \cdot \frac{1}{1/4}$. For figure 10 the median for every participant and difficulty level was taken and averaged over all participants. The resulting Total Task Time (TTT) is depicted in figure 10. As expected the times increase with a higher difficulty for both interaction styles while gaze interaction has a lower TTT than touch interaction up to a difficulty of 10. For higher difficulty levels gaze interaction became much more demanding for the drivers and most of them had to quit the task with difficulty 20. The interaction distracted the test persons too much from driving to continue the session in a reasonable manner. In
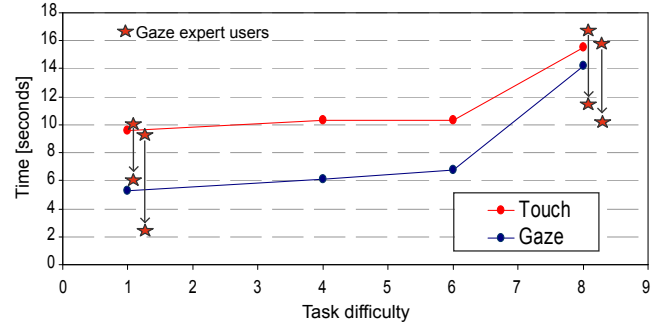

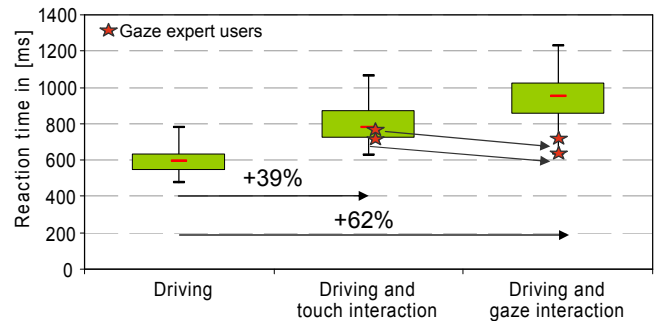Fig. 10. Total task times for different task difficulties

Fig. 11. Reaction times for both interaction methods compared to the baseline

average the TTT of the gaze system was ca. 31% lower than for the touch system. The two participants who already had previous experience with gaze interaction could even reduce their TTT by 38% which suggests that there is a steep learning curve for this kind of interaction.

Figure 11 shows a distinct increase in the median of reaction times of 39% for touch interaction compared to the baseline and 62% for gaze driven interaction what suggests that gaze interaction is more mentally demanding. However, for the gaze expert users the opposite tendency can be observed.

Also the percentage of acknowledged PDT dots decreases from a median of 100% to 79% for touch interaction and 73% for gaze interaction which confirms the findings from the reaction times. Again the gaze expert users performed better with the gaze interaction system than with touch interaction.

A measure to assess driving performance is the deviation of the vehicle center line from the lane center line. The higher the cognitive load of the secondary tasks (menu interaction) is, the less cognitive ressources are left for the primary driving task. As a result the deviation from the center line increases which can be observed in figure 13. The interquartile ranges of both interaction methods are higher and not overlapping with the baseline values. For both interaction methods these ranges are nearly equal but the maximum values differ strongly which suggests that some drivers are especially prone to distraction by gaze interaction. For the expert gaze users the tendency of a better performance with gaze interaction than with touch interaction can again be detected.

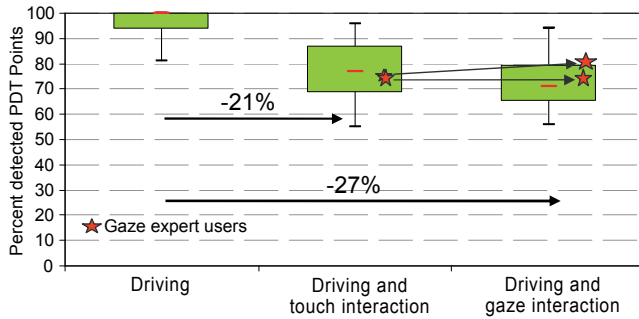An explanation of the constantly better performance of the

Fig. 12. Percentage of detected PDT point for both variants comapared to the baseline
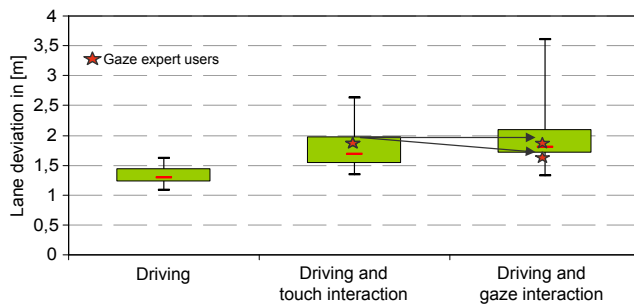

Fig. 13. Deviation of vehicle center line from lane center line in the Lane Change Task

trained gaze users could be that they use the system more intuitively. Untrained users tend to "stare" at the intended objects and focus all of their attention on this task. A trained user utilizes the advantage of the fast gaze selection method without thinking about the system processes behind their inputs. The result is a shorter TTT (see figure 10, upper stars denote TTT for touch, lower ones for gaze interaction) for trained users compared to untrained ones and a lower increase of TTT for higher task difficulty levels.

VI. CONCLUSION AND FUTURE WORK

We presented novel interaction methods for touch and gaze interaction on spatially distributed displays. Both interaction methods were evaluated in a simple lane changing scenario (Lane Change Task, see [9]) with identical tasks for gaze and touch interaction. The average outcome was a higher cognitive load for gaze interaction resulting in higher reaction times, less detected PDT points and lower performance in keeping the vehicle centered on the lane. For two gaze interaction expert users all results were better for gaze interaction compared to touch interaction.

Furthermore, a method to distribute various graphical widgets representing typical interface components of a state-of-the-art car dashboard on multiple displays is described. Portions of the main interaction screen represent the remote displays and widgets that are dragged onto these areas are moved to the respective displays. Interaction on the remote displays is realized using gaze-based multimodal interaction.

Future developments could include a functionality to keep the object on a display that was fixated last still selected. This would result in even shorter glances as the user does not have to keep one's gaze on an object until a value change is completed. Additionally speech feedback would be a promising feature to confirm the currently selected object while the user's gaze is already back on the road.

REFERENCES

[1] S. Bardins, T. Poitschke, and S. Kohlbecher, *Gaze-based Interaction in Various Environments*, Proc. 1st ACM International Workshop on Vision Networks for Behaviour Analysis VNBA, Vancouver, Canada, 2008.

[2] L. Harms, C. Patten, *Peripheral detection as a measure of driver distraction. A study of memory-based versus system-based navigation in a built-up area*, Transportation Research Part F: Traffic Psychology and Behaviour, 2003.

[3] A. Huckauf, M. Urbina, *Gazing with pEYEs: Towards a universal input for various applications*, Proceedings of the 2008 symposium on Eye tracking research & applications. Savannah, 2008.

[4] K. Eilers, F. Nachreiner, K. Hänecke, *Entwicklung und Überprüfung einer Skala zur Erfassung subjektiv erlebter Anstrengung*, Zeitschrift für Arbeitswissenschaft, 40, 1986.

[5] R. Jacob, *What you look at is what you get: eye movement - based interaction techniques*, Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people. Seattle, 1990.

[6] M. F. Land, *Eye movements and the control of actions in everyday life*, In Progress in Retinal and Eye Research, Volume 25, Issue 3, Pages 296-324, 2006.

[7] LC-Technologies, *Books written with the Eyegaze Communication System*, http://www.eyegaze.com/content/books-written-eyegazecommunication-system, 2009.

[8] K. Majlund Bach, M. Gregers Jæger, M. B. Skov, N. Gram Thomassen, *You Can Touch, but You Cant Look: Interacting with In-Vehicle Systems*, Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems. Florence, 2008.

[9] S. Mattes, *The Lane Change Task as a tool for driver distraction evaluation*, IHRA-ITS Workshop on Driving Simulator Scenarios, Dearborn, 2003.

[10] D. Miniotas, O. Špakov, I. Tugoy, I.S. MacKenzie, *Speech-Augmented Eye Gaze Interaction with Small Closely Spaced Targets*, Proceedings of the 2006 symposium on Eye tracking research & applications. San Diego, 2006.

[11] T. Poitschke, E. Bay, E. Drefers, H. Leinsle, M. Pavlic and G. Rigoll, *3D Gaze Tracking für die Mensch-Maschine-Interaktion*, Technical report, Lehrstuhl für Mensch-Maschine-Kommunikation, Technische Universität München, 2009.

[12] T. Poitschke, E. Bay, F. Laquai and G. Rigoll, *Using Liquid Lenses to Extend the Operating Range of a Remote Gaze Tracking System*, Proc. of IEEE Intern. Conf. on Systems, Man, and Cybernetics SMC 2009, San Antonio, Texas, USA, 2009.

[13] O. Špakov, D. Miniotas, *Gaze-Based Selection of Standard-Size Menu Items*, Proceedings of the 7th international conference on Multimodal interfaces. Trento, 2005.

[14] L. Thompson, M. Tönnis, C. Lange, H. Bubb, G. Klinker, *Effect of active cruise control design on glance behaviour and driving performance*, In Proceedings IEA 2006 16th World Congress on Ergonomics. Maastricht, 2006.

[15] C. Ware, H.H. Mikaelian, *An evaluation of an eye tracker as a device for computer input*, Proceedings of the SIGCHI/GI conference on Human factors in computing systems and graphics interface. Toronto, 1987.

[16] AttrakDiff Homepage: http://www.attrakdiff.de/en/AttrakDiff/Glossary/ accessed 30. March 2011.