Analyzing Gaze Behavior Prior to Interacting with a Multimedia Interface in a Car

Bastian Hinterleitner[1(✉)], Thomas Hammer[2], Stefan Mayer[1], Frederik Naujoks[2], and Nadja Schömig[2]

[1] Audi Electronics Venture GmbH, 85080 Gaimersheim, Germany
`bastian.hinterleitner@audi.de`
[2] Würzburger Institut für Verkehrswissenschaften GmbH, 97209 Veitshöchheim, Germany

Abstract. With the increasing number of functionalities of In-Vehicle Information Systems (IVIS) their complexity is rising. In addition new input modalities are used to control the car. Many manufacturers switch from using haptic input devices to touchscreens. On the other hand there are increasingly more sensors available in a car that could support the driver in their interaction with an IVIS by analyzing the driver's behavior. In this study gaze behavior prior to an interaction with was analyzed. Therefore 32 participants completed different tasks with an IVIS in a high-fidelity driving simulator. While driving, the participants were asked to navigate to a specific entry in a multi-level menu. Task and driving difficulty were modified as well as the input modality. One experimental group used a touchscreen and another one a rotary knob. Between the tasks, a short status information was presented to the participant either auditory or on a display. In 60.19% of the interactions, at least two preparatory glances were made. When using a rotary knob drivers had significantly less fixations on the touchscreen but those fixations lasted significantly longer. Difficulty of the driving task had no effect on the number of glances. When presenting information, the first glance is directed to the display the information appeared in. Using only sound and no visual information, the first glance is directed mostly to the center display rather than the cluster display. These findings should be considered for future design of IVIS and can help develop a more natural user interface.

Keywords: Gesture and eye–gaze based interaction · Metrics for HCI
Multimodal interface

1 Introduction

In-Vehicle Information Systems (IVIS) are available in virtually every car segment from luxury to low price vehicles. They mostly consist of a screen and some variation of rotary knob to control the interface. With the dropping prices of touchscreen displays, more and more manufacturers switch to touchscreen only systems without additional control elements. Whether this kind of interaction while driving is more distracting or not has been researched broadly. Jæger et al. [1] for example compared interaction via touch to gesture and buttons and found out that touch results in the longest glances away from the road. However, gaze behavior prior to interactions is mostly unknown. Gaze paths

and gaze distributions prior to an interaction with those systems may differ widely and should be considered when designing future IVIS. The time between the first glance at an interface and the actual interaction with it can possibly also be of interest for predicting user input. Especially in cars where both types of interaction are possible, the IVIS could then be able to predict which kind of control element the driver is going to use to appropriately support him. Therefore the differences in preparatory gaze behavior (i.e., just before the start of the interaction process) between using a touchscreen and a rotary knob were investigated in this study.

Different research questions arise when focusing on the gaze behavior prior to an interaction. First, we wanted to find out whether a gaze pattern can be identified when preparing for an interaction. Are there any glances on the interface the user wants to interact with before interacting? If so, is the user looking at the hardware control element he is using for the interaction or at the display on which the interaction is visualized? In addition we wanted to know how the difficulty of the driving task and experience with the system influences preparatory gaze behavior.

2   Related Work

In a prior study, we found out that gaze behavior alone is not sufficient enough to correctly predict single tasks but can be used to predict the location in a vehicle where the user is going to interact [2]. We used a head mounted eye tracking system (Tobii Pro Glasses 2) in a real car to analyze gaze paths of 20 participants while preparing the car for a journey. To ensure realistic gaze behavior, no additional tasks were given. We then used the gaze locations five seconds before an interaction as well as the type of interaction and, in a second step, the location of the interaction to train a machine learning algorithm. There were eleven different types of interaction such as "closing the door", "adjusting the seat", "using the navigation system" or "using the media controls" as well as two different locations ("center console", "steering wheel, seat and door"). Results for predicting the specific type of interaction were low (13.33%) but acceptable for the location of the interaction (70.00%) and could be improved by using an advanced algorithm. Concerning these findings a sufficient distance between areas of interaction needs to be considered for a prediction of the following interaction to succeed.

Tretten et al. [3] found out that there is a significant difference in gaze behavior depending on the location information is presented in the car. In their study 20 participants had to drive in a high-fidelity driving simulator through rural and urban driving situations with light and medium traffic. They then presented warnings in four different locations in the cockpit (head-up display, head-down display, infotainment display and center-stack display), which the participants had to respond. Response times and gaze patterns were measured. Half of the participants received warnings in the head-up and head-down display simultaneously while the other half received warnings in one of the four locations. Participants stated, that critical warnings should be placed in the head-up display and that the center-stack display is too far away to be looked at for warnings. There were significant differences in reaction times, number of glances, off-road gaze time and speed when comparing central location (head-up display and head-down

display) and peripheral location (infotainment display and center-stack display). However, sound was not used for presenting information in this study and gaze paths prior to reading the information were not analyzed.

To analyze gaze paths during interactions in a car usually gaze durations and transition probabilities for defined areas of interest are measured and calculated. This was done the first time in the early 1990s. Antin et al. [4] compared display based to paper based maps. They defined six areas of interest (roadway-centre, mirrors, instruments, roadway-off centre, moving map/paper map, signs/landmarks) and calculated the glance probabilities for each area and the transition probabilities between the areas. That way they were able to create a diagram for each tested system to visualize possible gaze paths. Dingus et al. [5] used the same method to test six different navigation conditions with a TravTek system, spoken navigation, a route description on paper and a conventional paper map. Those two studies were later compared by Dingus [6] with similar results. Those studies however always used all gaze data during one condition to create what they call "link diagrams". The probabilities for gaze paths before an interaction with the car in a relatively short timeframe were not analyzed.

3   Method

32 participants completed IVIS-tasks in a high-fidelity driving simulator. A real-world IVIS was transferred to the simulation environment. For interacting with the IVIS, a touch interface was used in one experimental condition, and in the other condition a rotary control knob in the center cluster. For different driving conditions, two variations of a car-follow-task as proposed by the NHTSA guidelines for evaluation of visual-manual in-vehicle tasks [7] were used simulating an easy and a difficult driving task. Experience with the system was compared over six tasks under each condition. The resulting gaze data was analyzed by looking at the amount, duration and path of fixations in the timeframe between each task instruction and the first interaction with the system (on average 6.60 s).

In addition, gaze behavior was analyzed when presenting status information to the driver at different positions and with different output modalities. These status messages were presented between the IVIS tasks and appeared either in the center display or in the cluster display. A third variation of those messages was presented using only speech and not displaying any information (see Table 1).

Table 1. Variations of status information presentation

Experimental condition	Visual indicator	Auditory indicator
#1	Text in the cluster display	Sound cue
#2	Text in the center display	Sound cue
#3	–	Spoken text

3.1 Setting

A static driving simulator with an Opel Insignia mockup was used. The front of the car was surrounded by a screen illuminated by three projectors providing a nearly 270° field of view. Additional displays were placed in the back of the car and on the side mirrors to allow for observation of the surrounding traffic. The driving simulator software SILAB was used.

During the study, the supervisor was sitting in an adjacent room separated from the simulator by a window. From that room he was able to monitor the test and see the participant through various cameras placed in the car. Additionally he was able to communicate with the participant using a microphone.

For gaze-tracking purposes a SmartEye System was used. This system consisted of four Basler infrared cameras and two infrared illumination units. The cameras were placed around the displays and the interaction elements used during the test to provide the best possible coverage of gaze-tracking availability (See Fig. 1).

Fig. 1. Simulator setup; left: rotary control knob; right: eyetracking cameras

In addition, several rgb and infrared based systems were used to record the behavior of the driver during the test.

3.2 Interface

For the interaction with the car's infotainment device a simple simulation of a current AUDI MMI GUI was implemented. It consisted of a hierarchical menu with different lists that the participants had to navigate through. The final selections in those menus did not trigger any actual action but gave a feedback depending on the task whether the selection was correct. In the top level or "main-menu" there were six different categories ("car", "sound", "radio", "media", "phone", "navigation") each leading to a separate list of entries to choose from. In each category a maximum number of four levels of menu depth was implemented and in each level there were few enough entry options so that the user was always able to see all possible entries of that level and hence did not have to scroll.

To provide similar interaction possibilities for both touch and a manual interaction knob, the center console of the car was modified. An AUDI MMI interaction unit was placed on top of the original Opel center console. This unit contained a rotary interaction knob and buttons for shortcuts for different main menu entries. The buttons around the touchscreen were modified to represent the same shortcuts.

3.3 Procedure

32 participants took part in the study. To achieve optimal availability of camera data, only persons with a minimal height of 1.64 m were invited. 19 men and 13 women participated with an age of 22 to 68 years (m = 32, sd = 10).

First, the participants had to sign a confidentiality agreement and were then presented with general information about the study. As all participants were familiar and trained in using this driving simulator beforehand no additional training for driving the simulator was conducted. To become familiar with the system, the participants were given six training tasks prior to the main study. Each task led to a different section of the menu so that every section had been explored before the actual test started.

Every participant completed the study in four different settings so that all combinations of track-difficulty (easy vs. difficult car following) and control element (touch screen vs. rotary knob) could be experienced. The sequence of those was randomized for all participants.

The driving task consisted of a variation of the car-follow-task as proposed by the NHTSA distraction guidelines. The participant was driving on a mostly straight road with two lanes. The task was to follow a car in front of the participant's car and keep a defined distance as constantly as possible. The distance between the two cars was displayed as a colored bar on the road in front of the participant. Depending on the distance the color of the bar changed. It appeared yellow when the distance was too large and blue when the distance was too small. Under perfect circumstances the bar was colored gray (within a range between 1.0 and 1.8 s time distance). Driving with the wrong distance however had no effect.

The car ahead changed its speed in a sinus-wave profile with slightly varying amplitude. That way the driver had to permanently adjust the speed of the simulated vehicle.

To create two levels of difficulty, the frequency of the driving profile was modified. In the difficult driving scenario the car changed its speed almost twice as fast as in the easy scenario.

Each participant had to do six tasks in each of the four settings. For example calling a friend or connecting their smartphone to the car's wifi. For starting the tasks defined positions on the route were used. If the participant wasn't able to complete a task in a defined driving distance, the task was cancelled to not overlap with the following one. The tasks were designed to be very similar and if possible to end on the fourth level of the menu so that the tasks didn't influence the gaze behavior differently. In addition, the task instructions (i.e., announcement about where to navigate in the menu) were presented auditory via prerecorded sound files so that the participant didn't need to look at a specific display before starting the task. If possible, all tasks ended on the deepest level of the menu. That means for each task, the participants had to go through at least

four levels of the menu. For some tasks the shortcut buttons around the rotary knob or the touchscreen could be used to switch to specific levels of the menu.

During each of the four test settings, two messages were presented to the driver to analyze gaze behavior when receiving information. Those messages were presented either spoken only or on one of the two displays (cluster or center display) and only between the menu tasks to not influence gaze data associated with them. The messages presented on a display were accompanied by a short sound cue so that the driver didn't miss them. Those messages were designed to inform the participant without the need to immediately interact with the car, for example: "low fuel, please visit a gas-station" or "low oil, please visit a garage".

After the test the participants received a financial compensation. The whole procedure took about 90 min.

3.4   Data Evaluation

Because the position of the rotary knob was very low in the vehicle setting several gaze samples didn't get recorded correctly or with insufficient quality when the driver was looking too far down. To be able to use and analyze those glances, it was decided to manually code all glances on a frame-by-frame basis using the SILAB video analysis tool. Therefore, every frame from the start of each task until ten seconds after the first interaction of the user has been analyzed and all glances towards the touchscreen and the rotary knob were determined. This didn't result in correct gaze directions but was sufficient to distinguish the different areas of interest needed for this study.

To determine whether there would be any preparatory gaze behavior prior to an interaction, the number of fixations on control elements was counted. This was done in the time between the instructions for the task and the first interaction with any control element. For this specific analysis, the type of control element was not distinguished. All fixations on the touchscreen and on the rotary knob were considered preparatory fixations.

The number of fixations on the different control elements were then compared for the different test settings. A Chi-square test was used to compare fixation-locations when using a touchscreen or a rotary knob and under different driving difficulties.

In addition, the mean durations of the fixations were compared for the two input modalities as well as the fixation durations after the first interaction. That way, differences in gaze behavior while interacting could be analyzed.

To analyze gaze paths prior to an interaction, the frequency of the gaze transitions between the three areas of interest (rotary-knob, touchscreen and windshield) was calculated. This was done for the four fixations prior to an interaction and only for the condition of using the rotary knob, as the touchscreen condition produced only gaze paths between the road and the touchscreen.

For the presentation of messages, the different fixations were counted and analyzed. The number of fixations off the road was counted and compared as well as the location of the first fixation. The time passed between the instructions and the first fixation as well as the duration of the first fixation were analyzed and compared as well.

4   Results

60.19% of the tasks showed at least two preparatory glances. Less than five percent of the participants showed five glances or more between presentation of the task and interacting with the car. Those glances usually target the display in which information is presented rather than the control element to manipulate this information. On average, the first fixation happened 3.70 s after the instructions (sd = 1.90 s) and 2.90 s before any interaction (sd = 2.61 s). When using a rotary knob, drivers had significantly less fixations on the touchscreen (p < .001) but those fixations lasted significantly longer (p < .001) (see Fig. 2). Difficulty of the driving task had no effect on the number of glances (p = .260).
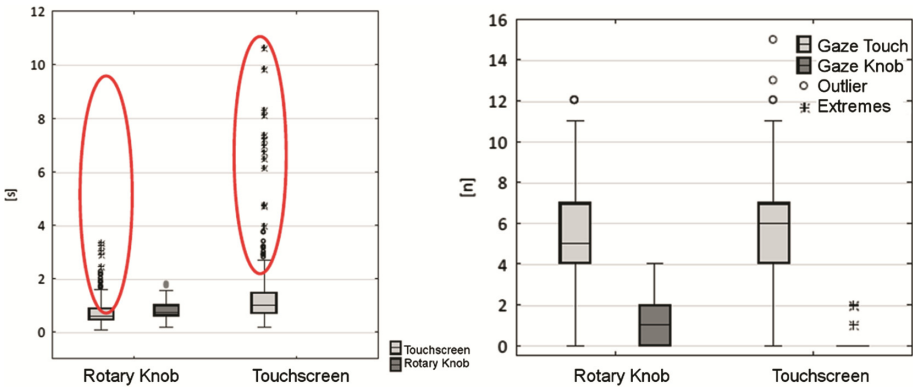

Fig. 2.   Differences between touchscreen and rotary knob; left: duration of the first fixation; right: number of fixations

When analyzing the path of fixations prior to the interactions, the study showed that the gaze of most users jumped between the road ahead and other areas of interest. There were very few transitions directly between displays or any display and the rotary knob. For creating a link diagram as proposed by Antin et al. [4] we used three areas of interest: road, touchscreen and rotary knob. Because of the lack of fixations on the rotary knob in the condition of using the touchscreen we only created the diagram for the condition of using the rotary knob. As in only 3.04% of all recorded interactions five or more glances were made prior to an interaction we used the first four glances between presentation of the task and the interaction with the car and created a link diagram for each (see Figs. 3 and 4). The first glance targeted in 60% of all conditions the touchscreen and in 30% the rotary knob which the participants had to interact with. Link diagrams were created for those two cases separately.
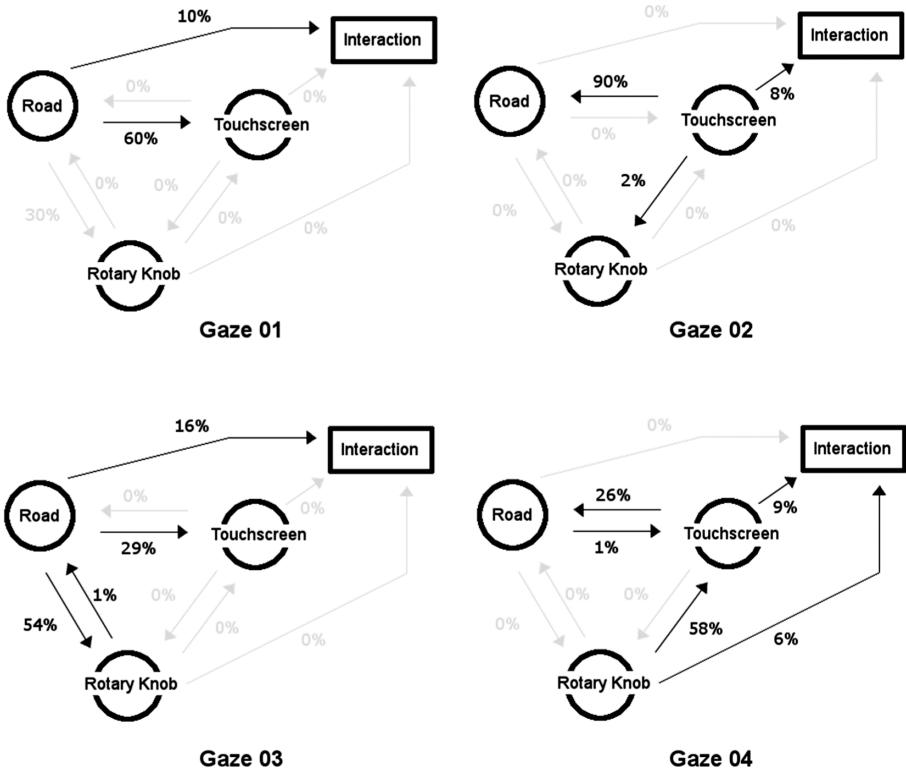
Fig. 3. Link diagrams for the first four glances prior to an interaction when the first glance is directed at the touchscreen

When looking at the touchscreen first the second glance was usually directed back to the road (90%) and the third to the rotary knob (54%) or to the touchscreen (29%) and then back to the touchscreen (58%) or the road (26%). We found out that in this case almost no transitions were made from the rotary knob to the road (1%) but usually back to the touchscreen first (see Fig. 3).

When the first glance was directed at the rotary knob the second glance almost always returned back to the road (97%). In 54% there was no third glance but the participant started interacting with the car. 36% of glances returned to the touchscreen, 10% to the rotary knob. With the fourth glance participants looked back to the road (69%) or started interacting (31%). According to these findings the most probable gaze path when looking at the rotary knob first is jumping back and forth between the road and the touchscreen or rotary knob (see Fig. 4).

Comparing the gaze paths when looking at the touchscreen or the rotary knob first it can be seen that the participants either scanned the display for information first then looked at the control element (rotary knob) and then back to the display or that they looked at the control element first then back to the road and then either interacted with the car right away or looked at the display. When looking at the rotary knob first
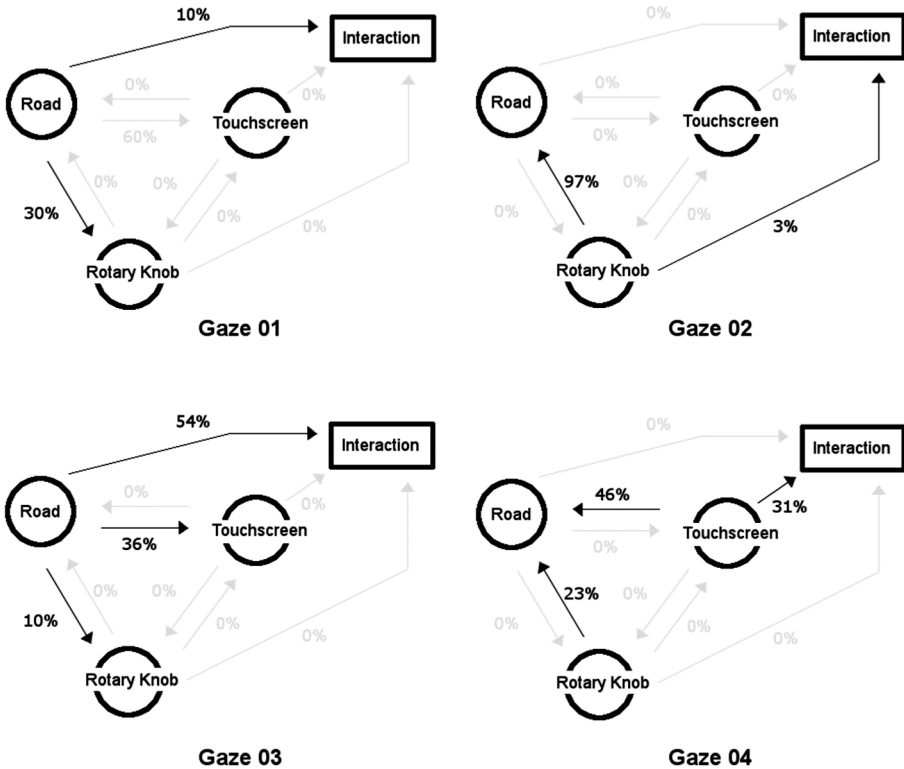
Fig. 4. Link diagrams for the first four glances prior to an interaction when the first glance is directed at the rotary knob

participants started to interact with the car earlier compared to looking at the touchscreen first. 54% of participants showing preparatory glances started interacting after the second glance when looking at the rotary knob first whereas only 16% started interacting after the second glance when looking at the touchscreen first.

When presenting status information to the driver his/her first glance is automatically directed to the display the information appeared in. Using only sound and no visual information the first glance is directed mostly to the center display rather than the cluster display (see Fig. 5). This may be due to the fact that information concerning the tasks was always presented in the center display in the rest of this study. The time of the first glance doesn't differ significantly between the three output variations (center display, m = 1.4 s, sd = 0.8 s; cluster display, m = 1.2 s, sd = 0.3 s; sound, m = 2.3 s, sd = 1.2 s), the deviation on the other hand is higher under the sound condition then under the other two. This might be due to the participant listening to the spoken message first before scanning the displays for written text. The time before scanning might vary strongly between participants.
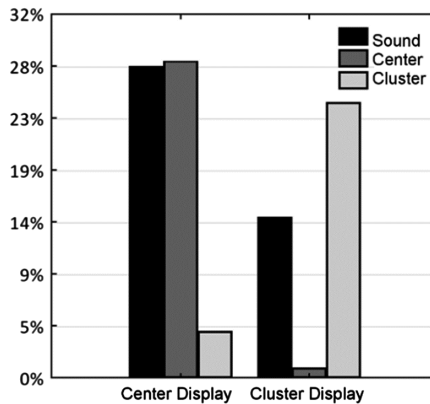
Fig. 5. Location of first glance when presenting information

5   Discussion and Implications

It could be shown that in most interaction scenarios there is a preparatory fixation of either the display or a control element with which the user wants to interact prior to the actual interaction. Concerning duration and number of glances there is a significant difference between using a touchscreen or a rotary knob. Also, there are almost no direct transitions between the single interaction elements. Instead, the driver fixates the road first before looking down again at another interaction element. When looking at a control element first the driver starts interacting after fewer glances then when looking at the display that control element interacts with first. These are important findings for the future design of in-vehicle infotainment systems. Especially for interfaces consisting of more than one display and control element it should be considered that a gaze path usually doesn't go from one element to the other but mostly across the windshield. This study also has implications for gaze distraction research as therefore the type of off-road glance before an interaction could support in identifying the corresponding task.

References

1. Jæger, M.G., Skov, M.B., Thomassen, N.G.: You can touch, but you can't look. In: Burnett, M. (ed.) The 26th Annual CHI Conference on Human Factors in Computing Systems, CHI 2008: Conference Proceedings, Florence, Italy, 5–10 April 2008, p. 1139. ACM, New York (2008)
2. Hinterleitner, B., Gauer, L.: Natürliche Blickfolgen vor einer Fahrt im Fahrzeug. Mensch und Computer 2017-Workshopband, pp. 441–449 (2017). https://doi.org/10.18420/muc2017-ws09-0305
3. Tretten, P., Normark, C.J., Gärling, A.: Warnings and placement positions in automobiles. In: World Congress on Ergonomics. Chinese Ergonomics Society (2009)
4. Antin, J.F., Dingus, T.A., Hulse, M.C., et al.: An evaluation of the effectiveness and efficiency of an automobile moving-map navigational display. Int. J. Man Mach. Stud. **33**(5), 581–594 (1990). https://doi.org/10.1016/s0020-7373(05)80054-9

5. Dingus, T.A., McGehee, D.V., Hulse, M.C., et al.: TravTek evaluation task C3-Camera car study (1995)
6. Dingus, T.A.: A meta-analysis of driver eye-scanning behavior while navigating. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 39, pp. 1127–1131. SAGE Publications, Los Angeles (1995)
7. National Highway Traffic Safety Administration: Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices (2012)