Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2010), November 11-12, 2010, Pittsburgh, Pennsylvania, USA

Making Use of Drivers' Glances onto the Screen for Explicit Gaze-Based Interaction

Dagmar Kern*, Angela Mahr+, Sandro Castronovo+, Albrecht Schmidt*, Christian Müller+

* Pervasive Computing and User Interface Engineering Group
University of Duisburg-Essen
Essen, Germany
{firstname.lastname}@uni-due.de

+ Automotive IUI Group
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
{firstname.lastname}@dfki.de

ABSTRACT
Interaction with communication and infotainment systems in the car is common while driving. Our research investigates modalities and techniques that enable interaction with interactive applications while driving without compromising safety. In this paper we present the results of an experiment where we use eye-gaze tracking in combination with a button on the steering wheel as explicit input substituting the interaction on the touch screen. This approach combines the advantages of direct interaction on visual displays without the drawbacks of touch screens. In particular the freedom of placement for the screen (even out of reach from the user) and that both hands remain on the steering wheel are the main advantages. The results show that this interaction modality is slightly slower and more distracting than a touch screen but it is significantly faster than automated speech interaction.

Categories and Subject Descriptors
H.5.2 [Information interfaces and presentation]: User Interfaces, User-centered design

Keywords
automotive, modality choice, timing, eye tracking

1. INTRODUCTION
Road safety is paramount for designing interactive systems in the car. It is widely agreed that new applications and devices should decrease the risks for drivers and their environment. Minimizing driver distraction is hereby a central issue and hence research often focuses on modalities that offer "hands-free" and "eyes-free" interaction. Design Guidelines for in-car user interfaces suggest that at least one hand has to remain on the steering wheel. There is a universal agreement that keeping both hands on the steering wheel is the safer way of driving and hence it is recommended to design system in a way that minimizes the time needed for manual interaction. In-car touch screens become more and more common as they allow a seamless adoption of well-known interaction paradigms (cash machine, mobile devices, ticket booth) for automotive applications. However, in order to use a touch screen in the car, the driver has to look at the screen, take off one hand from the wheel, and even lean forward in most cases. Thus, touch screens can neither be considered "hands-free" nor "eyes-free", nevertheless touch screens are in general considered "easy to use" even though reaching out to the screen can be strenuous and non-ergonomic. In the automotive context, touch screens have further disadvantages: first screens in large cars cannot be made touch-enabled as they are simply too far away - even more so when embedded into a deep-seated slot in order to reduce glare. And second due to the placement on the center stack, drivers, when pointing at an item on the screen, hide the area with their own hands impeding visual feedback.

The benefit of speech input as a well-known hands-free interaction technique is shown in many studies (see [1] for a brief literature review). However, the acceptance of speech input highly depends on its recognition rate. Other interaction forms that allow the driver to remain their hands on the steering wheel are researched like gesture-based input on the steering wheel [3].

In this paper, we propose an alternative approach with addressing the following question: Given that with touch-screen interaction the driver has to look onto the screen anyway, can we exploit this glance in order to avoid the need to take one hand off the wheel? In other words: can we use eye tracking as part of an explicit interaction?

Up to now, eye trackers are mainly used for driver distraction monitoring e.g. [2][10]. Recently, [7] introduced "Gazemarks" a new concept for facilitating attention switching. Here, the area on the screen is highlighted where users fixated last on the screen before moving away their attention. In this study, we propose and assessing a gaze-based approach that replaces the "touch" in touch-screen interaction by combing the gaze input on the screen with a button on the steering wheel.

2. PROPOSED GAZE-BASED INTERACTION
With the gaze-based interaction proposed, all interactions the driver is able to perform by single touch on a touch screen can be also performed by gazes. The principal design goal by designing this modality was that the duration of the gazes on the screen should not be longer than those necessarily occurring with the use of touch screens. Recent eye trackers with an update rate of 60Hz or 120Hz enable almost real-time feedback, so that the user only has to glance at an item. As opposed to touch, gaze interaction is indirect and needs means for providing feedback to the user as well as for selecting an item on the screen.

2.1 Highlighting an Item
Providing a feedback cursor for the eye is difficult as the calibration may not be perfect and as the human eyes are

AutomotiveUI'10, November 11-12, 2010, Pittsburgh, Pennsylvania
ACM 978-1-4503-0437-5.

permanently in at least slight motion. Therefore, we propose to provide visual feedback by highlighting the item the user looks at for example by framing the item, by changing its background color or by similar mechanisms that are used in systems using a turn-and-push dial. If not designed with a threshold this may make the highlighting toggle back and forth between different items on the screen. In order to avoid this effect, we are using a spatial and temporal threshold before switching to the next item. We move the visual feedback only if 5 gaze points have been registered at a new area. In our case where we have used an eye tracker with a data rate of 120Hz this means a delay of 0.04 seconds. The threshold time has to be as short as possible so that the user does not become aware of this delay and the size has to match the minimal size for which to give feedback (in our case a single word). An interaction sequence with a multimodal display is often interrupted as the driver should regularly look back at the street - again, our system is designed to not force more or longer gazes than touch screen. To address this issue, and ease the attention switching process, we keep the last item highlighted so that the driver can select it without looking back onto the screen (knowing that the correct item was in focus) respectively orientate faster when looking back, similar to the Gazemarks concept [7].

2.2 Selecting an Item.

There are two common ways to select an item with eye gazes: first by looking at an item in combination with pressing a button and second by a dwell time approach [5]. The dwell time approach requires the user to look at an item for a defined time period (about 150-250ms) for selecting it. Though [5] found that the dwell time approach is more convenient, we believe that button-pressed is more feasible in a driving condition because the driver is not forced to look at an object longer than necessary. Therefore, we propose a push-to-select button on the steering wheel (similar to the push-to-talk button).

3. APPLICATION DOMAIN

As mentioned earlier, the proposed gaze-based interaction approach can be applied in principle in any application domain where touch screens are used. Particularly, this can be menu item selection, list selection, selection item on a 2d space (such as points-of-interests on a map), grid selection (rows and columns of buttons). The example studied here belongs to the category of list selection. Particularly, we investigate the use of the proposed approach for error correction with speech input. Unconstraint dictation (for example for email or SMS) is the hardest form of automatic speech recognition (ASR) and remains error-prone, especially when the environment is suboptimal, because background noise is present or because the user can for some reason not fully concentrate on speaking like it is the case with driving.

Following the ASR post-correction paradigm [11, 6], we consider ASR as a black box, which cannot be optimized because we haven't got access to it. With dictating (and editing) text message while driving, the black-box scenario is very likely as car manufacturers are in the process of moving from owned on-board ASR solutions to off-board third party services such as Vlingo [13]. Obvious means for correcting errors that we know from personal computers at home will not be immediately applicable here, because in mobile situations a number of constraints apply. Moreover, correcting speech recognition error by adding another speech recognition step bears the danger of ending up in cascading error that frustrates the user [12]. We therefore believe that the problem of post-ASR correction is a suitable test-bed for the proposed gaze-based interaction concept.

4. EXPERIMENT

With experiments in the driving context, we speak of (simulated) driving as primary task, controlling vehicle functions (acceleration, steering) and other tasks related to driving as secondary task, and everything else (usually some sort of infotainment activity) as tertiary task. Following this scheme, we designed an experiment to compare touch screen (neither hands-free nor eyes-free) and speech (hands-free) interaction with the proposed hands-free gaze-based approach using post-ASR correction as a tertiary task.

Experiments of the use of interactive systems while driving are less constrained than typical desktop studies and hence we tried to minimize factors that impact the experiment. The following points need consideration. On the one side, touch screens are a well-established form of interaction with a simple technical realization that is familiar to the users in the study. On the other side, there is gaze-based interaction, which subjects are completely unfamiliar with and that additionally comes with a rather complex technical setup. Selecting a word-sized item on the screen is still at limits of the reliable spatial resolution of state-of-the-art systems. Considering that while we are primarily interested in "human factors", any - even minimal - technical changes will influence the respective condition. Similar problems were common in speech-based interaction a decade or more ago, when speech interaction was unfamiliar to most people. In ASR, the wizard-of-oz could be used in experiments to eliminate the impact of technical insufficiencies. Wizard-of-oz, however, cannot be applied to eye gaze, because reliably following the eye gazes of the subject manually in real-time and with high accuracy is not feasible.

Taking this issue into account, we formulate a conservative hypothesis: gaze-based interaction performs equally well with regard to speed and driver distraction as direct interaction with a touch screen while allowing the user to keep both hands on the steering wheel. We implemented a prototype for our gaze-based interaction approach and conducted an experiment with 24 participants.

Figure 1. Experimental setup (left). Exemplary correction sequence (right).

4.1 Stimuli

In order to measure effects on the critical part of the interaction and to be able to control the complexity and occurrences of errors and by this allowing for reproducibility of the results, we simulated a correction task in ASR. We presented a set of preprocessed sentences comprising of five to six words with eight

to nine syllables and measured the driving performance while subjects were correcting the error. Two third of the set were correction cases ("Peter has tree chocolate bars") and one third deletion cases ("You can see trf roadworks all around"). After a correction pass, the message was to be confirmed by activating a send button.

4.2 Apparatus and Performance Measures

Figure 1 illustrates the setup of our experiment. We placed the participants in a mock-up of a vehicle which is equipped with a game steering wheel, pedals, an Tobii™ eye tracker X120, an 8 inch touch screen, a microphone and a 42 inch display showing the driving environment.

We measured the driver distraction using the standardized "lane change task" (LCT) [8]. LCT was developed by Daimler and is currently in the process of becoming an ISO standardized tool (ISO Draft International Standard 26022). As stated in [8], it has already been successfully used in a large number of relevant studies. As a performance measure, LCT calculates the mean deviation of the lane between a normative model and the actual driving along the track. The performance of the baseline (only driving) is then compared with driving with a tertiary task in order to objectively assess the level of distraction induced by that activity.

Immediately after completion of the driving trial in the respective condition, the drivers were given a Driver Activity Load Index (DALI) questionnaire [9], derived from NASA TLX [4], assessing the subjective demands in the following standardized categories: 1) global attention demand: mental, visual and auditory demand required to complete the task; 2) visual demand only; 3) auditory demand only; 4) tactile demands: originally related to vibrations but here adapted to manual handling (there were no vibrations of any sort); 5) stress: fatigue, insecure feeling, irritation, discouragement, etc.; 6) temporal demand: pressure and specific stress felt due to timing; 7) interference: distraction of the driver induced by the secondary task. For each factor, the participants were asked to rate the level of demand felt during the session on a scale from 0 (low) to 5 (high) with regard to their usual driving. Each of the dali dimensions could be rated from 0 (low) to 5 (high) with regard to participants' driving in the baseline. As recommended in the standard procedure, we averaged the rating scores over all 7 dimensions as a global assessment of driving task workload.

The performance measure in the tertiary task was the number of sentences that could be corrected in a fixed period of time as well as the errors that were committed during that interaction.

4.3 Conditions

In the TOUCH condition subjects had to first touch the word to be corrected/erased. Then, a numbered list with three alternative words appeared with the fourth entry in the list being the word "delete". After selection a "send" button appeared at the upper right corner of the screen that had to be pressed to complete one interaction sequence before the next sentence was presented. With GAZE, subjects looked at the word to be corrected/erased. The respective word was framed. Instead of touching it on the screen, subjects had to press a designated button on the steering wheel. The rest of the interaction was designed analogously. Figure 1 illustrates a correction sequence. With SPEECH, each word in the original sentence was annotated with a superscripted number. The subject had to say the number of the word to be corrected/erased. Then, the subject had to say the number of the list entry: "1", "2", or "3" for one of the word alternatives and "4" for delete (the alternatives and the word "delete" were visible). In order to finish the interaction sequence, subject had to say the word "send".

4.4 Subjects and Procedure

For the experiment, 24 students have been recruited. The age of subjects ranged between 21 and 32 with an average of 25.2 years. The entire experiment took approximately 60 minutes. After introducing the eye tracker and calibration, a training drive was performed (without tertiary task), followed by BASELINE 1 drive. Then, the conditions TOUCH, SPEECH and GAZE were performed in a balanced layout (each condition including a short training time, LCT drive and a brief DALI questionnaire). LCT BASELINE 2 was determined afterwards. Finally, a summarizing questionnaire was presented.

4.5 Results

4.5.1 Lane Deviation

We determined the average deviation of the ideal line in meters for each condition/subject as described earlier. A repeated measures ANOVA was carried out with the following results: The main effect for condition was significant, ($F(4,20) = 29.7$, $p < .001$) indicating that driving performance was affected by the different conditions in the experiment. Pairwise comparisons (Bonferroni corrected) revealed no significant difference between the two BASELINE conditions (n.s.) which means that there is no considerable learning effect. In both BASELINE conditions, line deviation was significantly lower than in any of the experimental conditions ($p < .001$). Accordingly any of the tertiary task conditions turned out to decrease driving performance.

Condition	Mean	Std. Error	Avg.# Tasks completed
BASELINE 1	0.96	0.07	
BASELINE 2	0.89	0.06	
SPEECH	1.12	0.08	9.8
TOUCH	1.18	0.08	16.8
GAZE	1.31	0.11	14.1

Table 1: Deviation from the ideal line in the LCT for baselines and experimental conditions (mean and standard error); average number of tasks completed during each of the driving conditions.

SPEECH condition has the least deviation but it did differ significantly neither from TOUCH nor from GAZE (n.s.). However, when comparing the ideal line deviation of GAZE with TOUCH, the latter distracted significantly less from driving than the former ($p < .05$).

4.5.2 Subjective DALI ratings

Repeated Measures ANOVA revealed that there was as a significant difference between the conditions ($F(2,22) = 20.3$, $p < .001$). Helmert contrasts yielded a significantly lower overall demand for SPEECH than for TOUCH and GAZE

($F(1,23) = 41.9$, $p < .001$). The comparison of TOUCH and GAZE was not significant ($F(1,23) = .19$, n.s.).

4.5.3 Tertiary Tasks Completion and Errors

As the LCT driving task was invariably conducted on tracks with the same length and speed was constantly set to 60 km/h, the available time for completing the tasks was the same for every driver and condition. For task completion, a repeated measures ANOVA revealed a significant difference between the three conditions ($F(2,22) = 130.9$, $p < .001$). Helmert contrasts furthermore showed that in SPEECH significantly less tasks could be completed than in the other two conditions ($F(1,23) = 173.8$, $p < .001$). More tasks could be completed in TOUCH than in GAZE ($F(1,23) = 18.3$, $p < .001$).

Another aspect one should consider is the number of errors a subject committed in a specific condition. The number of errors was registered online by the experimenter. Accordingly, the number of errors committed in relation to the number of tasks fulfilled can be compared for the three conditions. A portion of the errors can be attributed to subjects making mistakes in the judgment which word should be corrected. However, a pretest revealed that the errors were easy to detect and evenly distributed across conditions, so this argument is rather outweighed. Also the assignment of sentences with respect to conditions was balanced. Hence, no difference in the difficulty of sentences could be held responsible for the differing error levels. The different technical states of the systems used for the conditions become most obvious here. All three conditions were compared by conducting a repeated measures ANOVA. A significant difference between the conditions could be measured ($F(2,22) = 8.1$, $p < .01$). Here GAZE differed significantly from SPEECH and TOUCH ($F(1,23) = 12.4$, $p < .01$) revealing that in GAZE more errors were committed than in the other two conditions. When comparing TOUCH and SPEECH subsequently, no significant difference in the amount of errors could be detected ($F(1,23) = .03$, n.s.).

5. CONCLUSION AND FUTURE WORK

In this paper we introduced a new modality for automotive user interfaces that uses eye-gaze tracking for explicit interaction in combination with a button. This interaction technique has interesting properties for car user interfaces as it can be operated while the hands remain on the steering wheel and the distance to the screen is not limited by the users reach. The experiment that we carried out showed that this modality is a valid alternative, even though it is a little slower than interaction on a traditional touch screen. In comparison to speech interaction using eye-gaze shows a speed benefit. Overall we conclude that this new modality is an alternative in cases where touch screens are not feasible and that it is a useful additional modality in combination with automated speech recognition. In future work we plan to assess how the interaction speed and influence on the distraction develops over time. From observation in the experiment we expect that users will become faster with the new modality and that after using it for some time it will be less distracting. Additionally we will explore how gaze in combination with a button on the steering wheel and speech input could become a general alternative to current touch screen interfaces.

6. ACKNOWLEDGMENTS

This work was funded by the German Ministry of Education and Research (project Car-Oriented Multimodal Interface Architecture, grant number 01IW08004).

7. REFERENCES

[1] A. Baron and P. Green. Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review. *Technical Report UMTRI 2006-5*, University of Michigan Transp. Res. Inst., 2006.

[2] L. Fletcher and A. Zelinsky. Driver Inattention Detection based on Eye Gaze--Road Event Correlation. In: *The International Journal of Robotics Research*, 28, 2009.

[3] I. Gonzalez, J. Wobbrock, D. Chau, A. Faulring, and B. Myers. Eyes on the road, hands on the wheel: thumb-based interaction techniques for input on steering wheels. In *Proc. of GI'07*, pages 95-102, 2007.

[4] S.G. Hart and L.E. Staveland. Development of a NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In: *Hancock PS, Meshkati N, editors. Human Mental Workload.* Amsterdam, Holland: 1988. pp. 139–83.

[5] R. J. Jacob. The Use of Eye Movements in Human-Computer Interaction Techniques - What You Look At is What You Get. *ACM Transactions on Information Systems*, 9(3), April 1991.

[6] M. Jeong, S. Jung, and G. G. Lee. Speech recognition error correction using maximum entropy language model. In *Proc. of INTERSPEECH* 2004, pages 2137-2140, 2004.

[7] D. Kern, P. Marshall, and A. Schmidt. Gazemarks: gaze-based visual placeholders to ease attention switching. In *Proc. of CHI 2010*, pages 2093-2102.

[8] S. Mattes. The lane change task as a tool for driver distraction evaluation. In H. Strasser, H. Rausch, and H. Bubb, editors, *Quality of Work and Products in Enterprises of the Future*. Ergonomia, 2003.

[9] A. Pauzié and G. Pachiaudi. Subjective evaluation of the mental workload in the driving context. *Traffic and Transport Psychology*, pages 173-182, 1997.

[10] B. Reimer, J. F. Coughlin, and B. Mehler, B. Development of a Driver Aware Vehicle for Monitoring, Managing & Motivating Older Operator Behavior. In *Proc. of the ITS-America,* Washington, DC. (2009).

[11] E. Ringger and J. F. Allen. A fertility channel model for post-correction of continuous speech recognition. In *Proc. of the 4th Intern. Conf. on Spoken Language* (ICSLP'96), pages 897-900, 1996.

[12] K. Vertanen and P. Kristensson. Parakeet: A continous speech recognition system for mobile touch-screen devices. In *Proc. of IUI 2009*, Sanibel Island, Florida, 2009.

[13] Vlingo Corporation. Vlingo mobile, 2010.