How Long *Can* a Driver (Safely) Glance at an Augmented-Reality Head-Up Display?

Nayara de Oliveira Faria, Joseph L. Gabbard
nfaria@vt.edu, jgabbard@vt.edu
Virginia Polytechnic Institute and State University

Augmented-Reality (AR) head-up display (HUD) is one of the promising solutions to reduce distraction potential while driving and performing secondary visual tasks; however, we currently don't know how to effectively evaluate interfaces in this area. In this study, we show that current visual distraction standards for evaluating in-vehicle displays may not be applicable for AR HUDs. We provide evidence that AR HUDs can afford longer glances with no decrement in driving performance. We propose that the selection of measurement methods for driver distraction research should be guided not only by the nature of the task under evaluation but also by the properties of the method itself.

INTRODUCTION

On the U.S. roadways in 2017, distraction-affected crashes accounted for 15% of injury crashes, 14% of police-reported motor vehicle accidents and 10% of fatal crashes (Highway Traffic Safety Administration & Department of Transportation, 2019). In the same year, 3,166 people were killed in motor vehicle accidents involving distracted drivers, many of which involved a secondary in-vehicle information system (IVIS), such as cellphones and GPS.

Augmented-Reality (AR) is one of the promising solutions to reduce distraction potential while driving and performing secondary visual tasks. By displaying graphical information directly on the windshield, within drivers' forward field of view, AR head-up displays (HUDs) allow drivers to use peripheral vision and their useful field of view and consume information without diverting attention away from the road. Several studies have shown the benefits of AR HUDs over traditional IVIS head-down displays in terms of better performance of vehicle control (Liu & Wen, 2004; Smith et al., 2016b), faster response time to an urgent event (Liu & Wen, 2004), and lower levels of mental workload (AblaBmeier et al., 2007; Liu & Wen, 2004).

Although AR HUDs' potential is promising, we currently don't know how to effectively evaluate interfaces in this area. With new AR HUDs capable of rendering images with larger fields of view and at varying depths, the visual and cognitive separation between graphical and real-world stimuli will be increasingly more difficult to quantify (Gabbard et al., 2019). Specifically, AR HUD interfaces exist within the line of sight needed to perform the primary visual driving task; and moreover, these AR interfaces may be present independent of whether or not drivers should be attending them.

Current practices for evaluating AR HUD user-interfaces are based on the Eye Glance measurement using the Driving Simulator test (EGDS) defined in the Visual-manual National Highway Traffic Safety Administration (NHTSA) Driver Distraction Guidelines (National Highway Traffic Safety Administration, 2012). This method was developed based on the tuning radio task empirical data, and it recommends that IVIS interfaces should be designed so that tasks can be completed by the driver while driving with glances away from the roadway of 2 seconds or less and total eyes off-road time (TEORT) of 12 seconds or less. Nevertheless, recent studies have shown that AR HUDs can afford longer glances with no decrement in driving performance (Smith et al., 2017, 2016a). As such, our studies to date suggest that current EGDS methods for assessing visual distraction for driving need to be readdressed and updated to account for AR HUDs. Specifically, to the scope of this study, there is no consensus on whether there is a threshold period for the safe execution of secondary tasks using AR HUDs. Therefore, we aim to answer: *How long can a driver safely glance at an AR HUD?*

Objectives

In this study, we induced a single, sustained glance that demands constant visual attention, such that the single glance duration is equivalent to total eyes-off-road time. We hypothesize that this approach will allow us to identify the glance duration (thus, consequently total eyes-off-road time) thresholds at which driving performance begins to wane. By carefully examining driving performance data across increasingly long glance duration, we aim *to establish several sustained glance duration thresholds for established dependent measures of driving performance.* We expect the longest glance durations below the identified thresholds to represent the best possible performance with the secondary task.

METHODS

Participants

Adapting NHTSA's driver distraction guidelines (National Highway Traffic Safety Administration, 2012) we recruited a sample of twenty-four gender-balanced licensed drivers with normal or corrected-to-normal vision. On average, participants were 22.36 years old (SD= 2.84) with 5.32 years of driving experience (SD= 2.63 years) and they reported to normally drive an average of 3923 miles per year. Due to driving simulator sickness, we had to remove data from two participants (one male and one female) and therefore, our final data sample consists of twenty-two participants. Three of the 22 participants drive a manual stick shift vehicle, while the rest drive an automatic vehicle. Nine participants reported driving a vehicle with a factory-fitted center console large screen display, one participant reported driving a vehicle with a large portable screen display, while the other twelve did not normally use such in-vehicle displays.

Equipment

We conducted this study in a fixed-base, medium-fidelity driving simulator in the Cogent Lab at Virginia Tech (Gabbard et al., 2019). This simulator is composed of the front half of a 2014 Mini Cooper cab fitted with a curved projection with 94 degrees of view, and both side and rear-view mirrors. The simulator also contains a 7" Lilliput USB monitor mounted directly behind the steering wheel to convey vehicle speed information. For this study, we equipped the simulator with a Pioneer Cyber Navi head-up display (HUD) with conformal AR graphics capabilities. The area displayed on HUD is 780x260 pixels, the field of view is 15 degrees and the virtual image position is approximately 3m away from the eyepoint. The driving simulator software was integrated with customized software, developed using X3D and Python, so that the AR HUD can provide real-time AR graphics perceptually overlaid into the dynamic CG-generated driving scene. That is, unlike other studies that render AR directly into a simulated environment (e.g. using virtual reality), our testbed renders AR graphics onto an aftermarket HUD, calibrated to a projected road scene to produce a more ecologically valid driver experience. During the study, participants wore SensoMotoric Instruments (SMI) eye-tracking glasses equipped with audio and video recording. We used iView Eye Tracking Glasses 2.6 software to collect binocular eye gaze data sampled at 60 Hz.

Experimental design

We used a 2x2 repeated-measures experiment such that each participant had a total of four drives. The presentation order of both *driving environment* (realistic, conventional) and *display type* (HUD, baseline) was counterbalanced using the Greco Latin-Square method so that potential order effects could be mitigated. The experimental conditions are a combination of the driving environments and display types; four levels described below:

- *Conventional HUD:* conventional environment, participants performed the secondary glance task
- *Conventional Baseline*: conventional environment, no secondary task performed
- *Realistic HUD:* realistic environment, participants performed the secondary glance task
- *Realistic Baseline:* realistic environment, no secondary task performed

For each drive, participants were exposed to four glance durations (20, 30, 40, 50 seconds), with three repetitions and randomized orders (see Figure 1 for experimental design overview).

Driving Task

Participants performed a car following task in which the lead vehicle driving behavior slightly changed according to the driving environment being used (realistic, conventional). There were no other additional vehicles on the same side of the road in which participants were driving and the lead car remained on the right side of the highway during the entire simulation. Participants were instructed to obey US driving laws and they started and ended each drive on the right-most lane of the road.

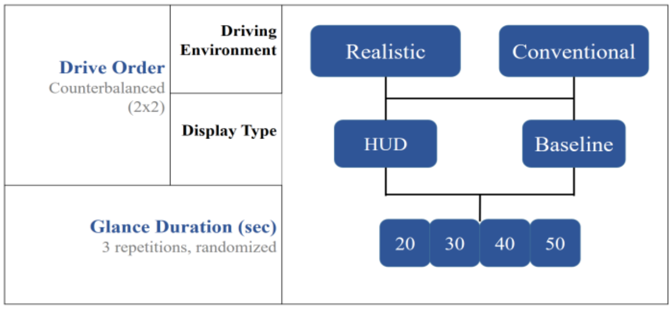

Figure 1: Overview of experimental design used in this study. Experimental condition is a combination of the driving environment and display type.

Conventional Environment

Our conventional environment was pursuant with NHTSA's driver distraction guidelines (National Highway Traffic Safety Administration, 2012) for driving simulator studies, which recommends that a lead car travels at a constant speed, on a straight road with no additional traffic. For this study, the speed was set at 55 mph and the participants were instructed to maintain a safe distance and not exceed the lead car in front of them.

Realistic Environment

Our realistic environment was designed to mimic a more realistic driving scenario in which the lead vehicle changes the speed at random times as participants are performing secondary tasks. In this scenario, the road included both straight portions and a slight curvature, with additional traffic traveling in the opposite direction. We included this condition because we were interested in investigating whether sudden changes in lead car speed and road curvature would impact the driving performance of the participants. For each glance task duration (20, 30, 40, 50), the road configuration was as follow: straight road with lead car breaking, curvature with no break and curvature with lead car breaking.

Required Glance Task

Participants performed a secondary required glance task concomitant with the primary driving task. This task was based on the random letter reveal technique derived from video production literature, whereby a single character randomly changes so quickly that participants cannot perceive any single character until the stimuli pauses and a target letter is revealed. In order to create sustained glances, we randomly revealed letters every 2-5 seconds with the letter target pauses between 0.5 and 1 second. These pauses were long enough to perceive the target but short enough to discourage participants from looking away. We also looked at eye-tracking videos to make sure that participants were not looking away when performing the secondary task. We also instructed participants to read the letter aloud during the paused to maintain engagement and establish a measure of secondary task accuracy.

Procedures

Our experimental protocol was reviewed and approved by Virginia Tech's ethics review board. Upon arrival at the lab, participants completed a short demographic survey using Qualtrics and consented to the research. Subsequently, they sat in the driving simulator, adjusted the seat to a comfortable driving position, and then we calibrated the HUD to ensure that all AR random letter graphics were perceived to be in the same location, regardless of participants' height and seat position.

Participants then performed a practice drive to get acquainted with driving simulator functionalities, car dynamics and secondary task to be performed. We explained to participants the letter reveal task and instructed them to read the target letter aloud whenever they perceived a pause. This familiarization drive lasted at least five minutes and ended when we checked that the participant correctly understood the letter reveal task and was comfortable and able to keep vehicle control while starting, stopping, turning and driving straight.

Before each drive, we administered a driving simulator sickness questionnaire to monitor participants' levels of discomfort using the simulator. When participants rated high levels of simulator sickness, we encouraged them to take a break and asked whether they felt comfortable continuing the study. If a participant decided to withdraw from the study due to simulator sickness, $15 was compensated for their time and their driving data was not used for further analysis.

Each data collection drive lasted between 12 and 15 minutes. After each drive, participants completed the NASA Task Load Index (TLX) questionnaire to record their perceived workload. After completion of the study, individuals signed a post-trial consent form and were compensated $15 for their time

ANALYSIS

We performed a mixed-effects ANOVA linear model to account for individual participant differences as a random effect. We measured subjective workload using the NASA TLX questionnaire in which the average of sub-scale scores comprised the raw TLX score. Lateral vehicle control was analyzed in terms of the standard deviation of lane position (SDLP) and, longitudinal vehicle control was analyzed in terms of the standard deviation of speed and average headway. We only evaluated driving performance data corresponding to the time during which participants performed secondary tasks. For each dependent measure, we used Minitab 19 and JMP Pro 15 to fit our model, accounting for the effect of the independent measures (glance duration, condition, order, participant, and gender) and second-order interactions of these effects.

RESULTS

NASA Task Load Index (NASA-TLX)

We found a significant effect of condition on overall raw TLX score ($p<0.000$), mental demand ($p<0.000$), physical demand ($p<0.021$), temporal demand ($p<0.000$), effort ($p<0.000$), frustration ($p<0.000$), and distraction ($p<0.000$). Post-hoc Bonferroni analysis showed that overall participants felt that performing secondary tasks in the realistic environment is more distracting and more cognitive demanding than performing the same tasks in a conventional environment (see Figure 2)

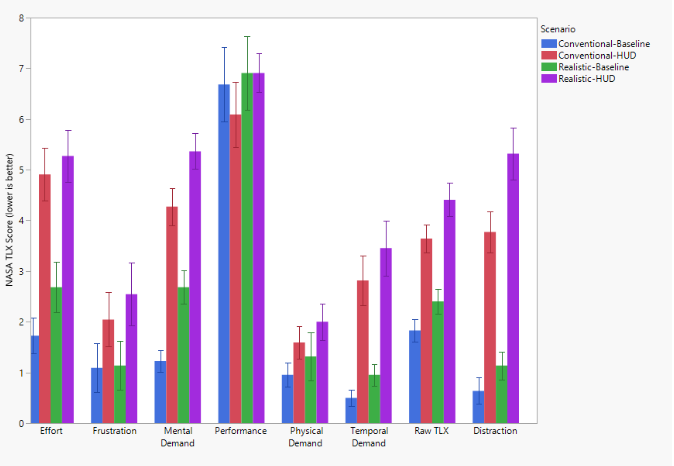

Figure 2: Participants rated the NASA TLX sub-scored on a scale of 0 (low demand) to 10 (high demand). Each error bar is constructed using 1 SEM.

Driving Performance

Standard Deviation of Lane Position (SDLP)

We found a main effect of duration ($F_{(3,\ 1019)} = 4.77$, $p<0.003$) and condition ($F_{(3,\ 1019)} = 11.44$, $p<0.000$) on the SDLP. Table 1 and Table 2 present Bonferroni post-hoc results for these main effects in which means do not share a letter are significantly different.

Table 1-SDLP Bonferroni post-hoc for duration

Duration	Mean	Grouping	
50	0.962599	A	
40	0.912620	A	B
30	0.885124	A	B
20	0.829595		B

Table 2-SDLP Bonferroni post-hoc for condition

Condition	Mean	Grouping	
Conventional-Baseline	0.976088	A	
Realistic-Baseline	0.962219	A	
Realistic-HUD	0.853381		B
Conventional-HUD	0.798251		B

Standard Deviation of Speed

We found a main effect of condition ($F_{(3,\ 1019)} = 4.77$, $p<0.003$) on the standard deviation of speed. Table 3 present Bonferroni post-hoc results for this main effect in which means do not share a letter are significantly different. Figure 3 presents a graphical representation of this main effect by the duration of the secondary task.

Table 3-Stde of Speed Bonferroni post-hoc for condition

Condition	N	Mean	Grouping		
Realistic-HUD	264	8.92584	A		
Realistic-Baseline	264	8.02900		B	
Conventional-HUD	264	2.41958			C

Conventional-Baseline	264	1.54402	D

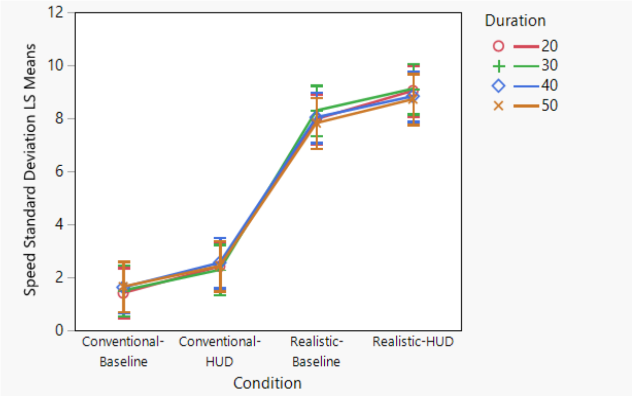

Figure 3: Least Square means plot for speed standard deviation

Average Headway

We found a main effect of condition ($F_{(3,\ 1027)} = 13.21$ $p<0.000$) on the average headway. Table 4 present Bonferroni post-hoc results for this main effect in which means do not share a letter are significantly different.

Table 4-Average headway Bonferroni post-hoc for condition

Condition	N	Mean	Grouping	
Realistic-HUD	264	197.634	A	
Conventional-HUD	264	193.849	A	
Realistic-Baseline	264	169.332		B
Conventional-Baseline	264	163.464		B

DISCUSSION

This study provides an initial examiniation of potential measures for evaluating the impact of AR HUDs on driving performance, and in turn to inform the safe design, deployment, and adoption of AR technology in vehicles. Our major goal was to establish the longest possible safe glance as a baseline that would serve subsequent methods development, given that current standards presume a 2.0 second glance duration and 12.0 seconds total eyes off road time (TEORT) as the threshold for safe driving (Driver Focus-Telematics Working Group, 2006). In this research, therefore, we induced prolonged sustained glances (20, 30, 40 and 50 seconds) which we presumed corresponded to the total-eyes-off road time (because the participants needed to deviate attention and gaze concentration from the road to the secondary task being performed). We investigated driving performance in terms of the standard deviation of lane position, the standard deviation of speed, and average headway. Below, we discuss the importance and relevance of our results.

Driving Performance

SDLP was found to be higher for a glance of 50 seconds compared to a glance of 20 seconds, and this difference is statistically significant. We suggest that the upper threshold standard of 12 seconds may not apply for AR HUDs, as TEORT of 30 and 40 seconds did not show degradation of any other driving performance measures. In fact, only SDLP performance started to wane when sustained glance duration reached 50 seconds. Because we did not include road hazards, and we instructed participants to travel at a safe distance from the lead car, we believe that these factors contributed to no statistical differences in terms of longitudinal vehicle control. Additionally, we did not measure drivers' ability to detect and react to events when performing long secondary tasks using AR HUDs. We believe that event detection performance will wane more quickly than lane-keep performance, and so, the upper threshold in which drivers can safely perform AR HUDs tasks could be even lower. Also, we only analyzed our data in terms of average time-frames corresponding to each sustained glance. As a result, extreme range behaviors are not well-described using this method of analysis. Finally, longitudinal and lateral vehicle control measures might interact not only with task duration but also with hazards detection and task difficulty. Such considerations should, therefore, be further investigated in future studies.

Driving Environment

In this study, we investigated whether a more realistic driving scenario would influence driving performance when compared to NHTSA-prescribed conventional scenario for driving simulator user-studies. As mentioned, because AR HUDs present information in front of the driver's field of view, we believe simplistic conventional environments used to assess HUDs may under-emphasize their actual impact on drivers. Even though in the realistic environment we have added more cognitive load and work by asking people to respond to the curves in the road and different speeds, we believe that this practice gives us a better understanding of what people are actually capable of when using HUDs in real-life settings. Also, AR per se imposes to the driver perceptual challenges that are not present in traditional head-down displays (e.g., AR can occlude the real-world). Even the most parsimonious AR graphic has the potential to block and obscure important elements of the driving scene. Therefore, a more realistic representation of real-life environments should be used when evaluating HUDs. We strongly believe that although drivers may be able to compensate for their driving performance when the driving scenario is relatively easy, they will have far less capacity to do so in more complex and challenging driving situations. As we have shown, methods relying on conventional driving environments may not highlight differences in driver behaviors due to other environmental factors.

Reference Task and Performance Criteria

In order to analyze the impacts of in-vehicle displays on driving performance, a reference task - task used as a reference point for determining the maximum level of secondary demands that are considered acceptable to the driver (Regan et al., 2009) - is usually employed. In this case, if driving performance while performing a secondary task is poorer than the baseline reference task, then the secondary task being evaluated is assumed to be unsafe to perform while driving. In this study, our approach is to assume that the reference task is no task (baseline condition).

Regarding the average headway, drivers determined their own headway – their only instructions were to follow at a “safe” distance. For this measure, we have found that participants adopted longer distance between their own vehicle and the lead car (statistically similar in both driving environments) when interacting with the secondary letter reveal task as compared to the baseline. Similar results have been found in the literature, suggesting that drivers tend to adopt longer headways behavior when performing visual tasks concomitant with driving in order to maintain safety margin (Greenberg et al., 2003; Östlund et al., 2004).

Surprisingly, the performance of the SDLP was worst in both conventional and realistic environments during the baseline drive. There are two plausible explanations for this phenomenon: *cognitive load* and *gaze concentration effect*. The letter reveal task increased the cognitive load required from participants – the amount of cognitive resources required to perform a competing activity while driving - and so, SDLP may have improved. Several studies have found the same pattern in which the presence of moderate cognitive load enhanced lane-keep performance while driving as compared to a baseline condition (Engström et al., 2005). In addition, we have induced sustained glances towards the center of the road (see Figure 4), and therefore, we induced the gaze concentration effect, which was also found to improve lateral vehicle control as compared to baseline conditions (Engström et al., 2005; Reimer et al., 2012). It is important to emphasize that the gaze concentration effect is not the same as tunneling vision, and, that the latter was not investigated in the scope of the present study.

Figure 4: Representation of the secondary glance task paused on the letter H. Red circle represents a possible glance concentration effect.

Because HUDs afford superimposition of information onto the driving road, we need to be more careful when choosing appropriate methods to evaluate their effect on driving performance. We have shown that due to the HUD location of our secondary task, the gaze concentration effect was most likely induced in our participants, and therefore, the reference task used is not the best approach to be chosen.

CONCLUSION

How long can a driver (safely) look at an AR HUD? In this study, we have shown that current visual distraction standards for evaluating in-vehicle displays may not be applicable for AR HUDs. We propose that the selection of measurement methods for driver distraction research should be guided not only by the nature of the task under evaluation but also by the properties of the method itself. There are still a number of human perceptual factors that should be taken into consideration when designing sensitive, reliable and valid methods for evaluating AR HUD effects on human performance. In future studies, we will expand the scope our of research by broadening the concept of driving performance to take into account drivers' ability to detect events (e.g. inatentional blindness and change blindness) when performing long AR HUD tasks.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 1816721. We'd also like to thank Kyle Tanous for the technical support in creating the AR HUD stimulus.

REFERENCES

AblaBmeier, M., Poitschke, T., Bengler, F. W. K., & Rigoll, G. (2007). Eye Gaze Studies Comparing Head-Up and Head-Down Displays in Vehicles. *Multimedia and Expo, 2007 IEEE International Conference On*, 2250–2252. https://doi.org/10.1109/ICME.2007.4285134

Driver Focus-Telematics Working Group. (2006). *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems*.

Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(2 SPEC. ISS.), 97–120. https://doi.org/10.1016/j.trf.2005.04.012

Gabbard, J. L., Smith, M., Tanous, K., Kim, H., & Jonas, B. (2019). AR DriveSim: An Immersive Driving Simulator for Augmented Reality Head-Up Display Research. *Frontiers in Robotics and AI*, *6*, 98. https://doi.org/10.3389/frobt.2019.00098

Greenberg, J., Tijerina, L., Curry, R., Artz, B., Cathey, L., Grant, P., Kochhar, D., Kozak, K., & Blommer, M. (2003). EVALUATION OF DRIVER DISTRACTION USING AN EVENT DETECTION PARADIGM. *Journal of the Transportation Research Board*, *1843*, 1–9.

Highway Traffic Safety Administration, N., & Department of Transportation, U. (2019). *Research Note: Distracted Driving in Fatal Crashes, 2016*.

Liu, Y.-C., & Wen, M.-H. (2004). Comparison of head-up display (HUD) vs. head-down display (HDD): driving performance of commercial vehicle operators in Taiwan. *International Journal of Human-Computer Studies*, *61*(5), 679–697. https://doi.org/10.1016/J.IJHCS.2004.06.002

National Highway Traffic Safety Administration. (2012). *Visual-Manual NHTSA Driver Distraction Guidelines For In-Vehicle Electronic Devices*. https://doi.org/10.1558/jsrnc.v4il.24

Östlund, J., Nilsson, L., Carsten, O., Merat, N., Jamson, H., Jamson, S., Mouta, S., Carvalhais, J., Santos, J., Anttila, V., Sandberg, H., Luoma, J., De Waard, D., Brookhuis, K., Johansson, E., Engström, J., Victor, T., Harbluk, J., Janssen, W., … Burns, P. (2004). *Title HASTE Deliverable 2 HMI and Safety-Related Driver Performance*.

Regan, M. A., Lee, J. D., & Young, K. L. (2009). *Driver Distraction: Theory, Effects and Mitigation*. CRC Press.

Reimer, B., Mehler, B., Wang, Y., & Coughlin, J. F. (2012). A field study on the impact of variations in shortterm memory demands on drivers' visual attention and driving performance across three age groups. *Human Factors*, *54*(3), 454–468. https://doi.org/10.1177/0018720812437274

Smith, M., Gabbard, J. L., Burnett, G., & Doutcheva, N. (2017). The effects of augmented reality head-up displays on drivers' eye scan patterns, performance, and perceptions. *International Journal of Mobile Human Computer Interaction*, *9*(2), 1–17. https://doi.org/10.4018/IJMHCI.2017040101

Smith, M., Gabbard, J. L., & Conley, C. (2016a). Extending Measures and Metrics of Visual Distraction with Automotive Head-up Display Applications. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - Automotive'UI 16*, 185–192.

Smith, M., Gabbard, J. L., & Conley, C. (2016b). Head-Up vs. Head-Down Displays. *Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - Automotive'UI 16*, 185–192. https://doi.org/10.1145/3003715.3005419