Automotive Research in the Public Space – Towards Deployment-Based Prototypes For Real Users

Michael Braun
BMW Group Research, New Technologies, Innovations
Garching, Germany
LMU Munich
Munich, Germany
michael.bf.braun@bmw.de

Florian Alt
Bundeswehr University
Munich, Germany
LMU Munich
Munich, Germany
florian.alt@unibw.de

Florian Roider
BMW Group Research, New Technologies, Innovations
Garching, Germany
University of Bamberg
Bamberg, Germany
florian.roider@bmw.de

Tom Gross
University of Bamberg
Bamberg, Germany
tom.gross@uni-bamberg.de

AutomotiveUI '18 Adjunct,, September 23–25, 2018, Toronto, ON, Canada
ACM 978-1-4503-5947-4/18/09.
https://doi.org/10.1145/3239092.3265964

Abstract
Many automotive user studies allow users to experience and evaluate interactive concepts. They are however often limited to small and specific groups of participants, such as students or experts. This might limit the generalizability of results for future users. A possible solution is to allow a large group of unbiased users to actively experience an interactive prototype and generate new ideas, but there is little experience about the realization and benefits of such an approach. We placed an interactive prototype in a public space and gathered objective and subjective data from 693 participants over the course of three months. We found a high variance in data quality and identified resulting restrictions for suitable research questions. This results in concrete requirements to hardware, software, and analytics, e.g. the need for assessing data quality, and give examples how this approach lets users explore a system and give first-contact feedback which differentiates highly from common in-depth expert analyses.

Author Keywords
Automotive UI; User Studies; Deployment; Prototypes

CCS Concepts
•Human-centered computing → HCI design and evaluation methods;

Motivation

User studies follow the primary goal to get feedback from a subset of users which is representative of the future demographics of a product. HCI researchers often struggle to fulfil the requirements for a suitable user sample which has been more and more criticized recently and is one of the contributors to the problem of insufficient replicability of many HCI studies [3]. In the automotive UI community, we see studies with students and employees as participants. This makes sense on one hand as they are easier to recruit and have the neccessary NDAs in place. On the other hand there might be research questions which should better be answered by real users without prior knowledge. Another problem are the sample sizes we are working with. The mean sample size from the 28 driving studies published at AutoUI 2017 was 24.4 ± 10.5 with a slight imbalance towards male participants (57%). The mean age was 29.9 ± 5.9 years deviating 8.3 ± 4.5 years. While this might be sufficient for qualitative research and within-subject designs, such sample sets will seldom be representative of the core target group and might lead to unreliable statements when approached with quantitative methods [5].

Related Work

A possible approach to tackle the problems of small sample sizes and limited demographic viability are deployment-based user studies. We can learn from other domains such as public displays [1] and mobile systems [4] how large numbers of actual users can be incorporated into a study setup in order to get a more realistic user sample and more measuring points as input. Alt et al. provide an overview of research approaches for public displays, stating system evaluations in a public setting can be difficult due to automation and legal issues, yet they offer the advantage of high ecologic data validity [1]. Henze et al. question the practice of data collection in lab studies as deployments can generate a multiple of quantitative data points [4]. Deployment-based research for automotive user interfaces, meaning the installation of unreleased in-vehicle information systems (IVIS) in series vehicles is however problematic from a legal perspective, expensive and potentially dangerous for participants.

We introduce an approach that aims at collecting valid user data by incorporating a large and diverse group of participants. Therefore, we deployed an interactive automotive prototype with a driving simulation in a public space. We describe requirements and lessons learned regarding the procedure, the prototype and the collected data.

Requirements

In order to build a useful prototype, we first need to identify the requirements for such a system. They can be partially taken over from real-world driving studies (e.g. [2]), but also depend on the intended purpose of displaying the simulator publicly.

Clarity

As participants can freely partake in the experiments, all instructions need to be clear and comprehensive, even without an experimenter explaining the procedure.

Presentation

Participants are not recruited through usual channels, but decide to participate spontaneously based on the prototype's attractiveness. This means that the booth needs to generate a certain pull effect for passers-by. An anonymous packaging concept should prevent biasing user ratings with brand expectations. We also have to take into account that people will often arrive in groups, so we can use potential bystanders for qualitative feedback on ideas and at the same time avoid distraction for the study participant.

Figure 1: The seating box combines a driving simulation with sensors (gaze, gestures, video) and a versatile combination of touch and ultra wide screen display to allow prototyping for future IVIS.

Figure 2: Prototype situated in a public space. Groups of passers-by often felt intrigued by the driving simulator and consecutively took part in the experiment.

System Stability
The software prototype must be robust to all kinds of (unintended) user input. In case of inconsistent software states it must be easy to reset the system ("Reset"-button). In the worst case, a reboot of the entire system should reset all systems into a working state. This is especially important for local staff who might not have a technical background.

Adaptivity
Participants come in all ages and sizes, resulting in different ergonomic requirements such as adaptable seat and pedal positions. There is also a great variance in the levels of prior experience with digital systems. The procedure needs to be understandable and feasible for everybody.

Data Collection
Any form of data acquisition or input that is usually controlled by the experimenter has to be automated, e.g. the calibration of eye-tracking sensors. There has to be additional data to retrace and assess the success of such automated procedures. On the other hand, data must be anonymized in conformity with data protection regulation and only necessary data should be saved.
Qualitative feedback from users needs to be collected within the study procedure as well. We need to embed this functionality in the system and find out how much time participants are willing to spend on questionnaires in order to optimize efficiency.

Remote Access
Off-site experimenters need to have remote access to the data and software. This way they can check quality of the data, identify upcoming issues and make software updates over the air.

Implementation
The hardware components of the seating box consist of a Microsoft Surface tablet as main UI station and for touch input, a 34" LG curved screen for the dashboard with a webcam mounted on top, a Logitech steering wheel and pedals, a Tobii 4C eye-tracker, a Leap Motion gesture sensor and an adjustable driving seat taken from a decommissioned BMW i3. All components are enclosed in a wooden frame we had built to resemble the ergonomic design of a middle class vehicle. The curved screen is mounted flexible in height behind a partial cover, making it appear as an ultra wide screen display with 86 cm width and 8-14 cm height (see Figure 1). The driving simulation is built in Unity 3D and runs on an MSI Nightblade attached to a 70" wide screen display. Both computers are controlled by wireless keyboards stowed under the driver arm rest. The study procedure was as well implemented with Unity 3D.

We planned the procedure meticulously as every usage error would result in data loss or inaccuracies. Textual instructions on the tablet are communicated step by step and in easy words. Participants had to press a "Continue" button after each step so they can read at their own pace.

Data acquisition was automated from the point participants press start. A hidden trial button allows for test rides by the staff which are marked in the log files as such. An initial questionnaire queries participants' demographic data and exclusion criteria like inebriation. In the first deployment we collected gaze and gesture data and we analyzed the video feed for facial expressions. This data was continuously logged, the video stream itself however was never saved to disk to maintain anonymity. After the ride participants were asked to rate a total number of 23 likert-scale items on the tablet.

Figure 3: A participant interacts with the prototype. In this study users selected items using hand gestures.

The booth was designed as a white label test without recognizable company affiliation. We provided general information on research questions and explained the technology behind our prototype on an information screen. An open feedback wall allowed bystanders to give feedback on the ideas presented. We also provided co-creation sheets where users could give their preferences on gesture interaction and give design recommendations for future IVIS.

Lessons Learned

The driving simulator worked very well in attracting passers-by, this was most probably also connected with the futuristic look and the location in a highly frequented city center. We learned that it is hard to control experiment completion without supervision: participants often left during the ride or when they were bored by the questionnaires. We also refined small aspects of the procedure, e.g. adding a notification sound when user action was required. Here, the possibility to deploy software updates over the air was very helpful. It allowed us to iterate on wording after we recognized that some of the users could not follow the textual instructions very well. Multimedia content, such as images or videos could provide a better understanding for participants.

The hardware had to endure constant stress, which showed occasionally during the testing period. Several sensors had to be replaced as they broke from running hot. The steering wheel was dismounted twice and finally broke under the weight of users who used it to pull themselves out of the seat. We repaired the wheel and added warning stickers which helped prevent further damage to the wheel. We learned that common gaming hardware which is often used in lab setups does not withstand the physical requirements of studies with several hundred participants. In future deployments we need to focus even more on the sturdiness of the used hardware.

Of the 693 participants in our initial 3 month study, 435 finished the ride and questionnaires. Of those 39 were under the age of 17, 11 said they were intoxicated. Another 56 only clicked through the questionnaires which we detected by calculating the standard deviation of button presses. This results in 329 valid sets of qualitative data (47.3%). Concerning gaze data, we report successful calibration in 91 cases (13.1%). The valid sample group was aged 36.4 ± 14 years and consisted of 109 women and 219 men.

Feedback from the co-creation sheets was overwhelming with 497 participants (298 male, 199 female) with a mean age of 32 years. The open feedback walls provided some good input but also a lot of superficial comments and opinions.

Conclusion

User studies in deployment environments offer the advantage of a more heterogenous sample group than lab studies, yet they come with limitations. As the setting is not as controllable, they are best suited for the collection of qualitative data, which we also know from real-world driving studies [2]. The study procedure needs to be appealing and fun for participants, as they are free to leave when they feel to. The requirements to the robustness of the setup for hardware and software go beyond the requirements of user studies in lab settings. Researchers have to prepare hardware and software accordingly, but also consider means for quick intervention in case of failure, such as over the air updates. The data collected in this type of experiment also has to be filtered profusely, resulting in a lot of waste. The acquired qualitative data was a great asset to work with, quantitative approaches might however not see a major benefit compared to conventional lab studies.

REFERENCES

1. Florian Alt, Stefan Schneegaß, Albrecht Schmidt, Jörg Müller, and Nemanja Memarovic. 2012. How to Evaluate Public Displays. In *Proceedings of the 2012 International Symposium on Pervasive Displays (PerDis '12)*. ACM, New York, NY, USA, Article 17, 6 pages. DOI: http://dx.doi.org/10.1145/2307798.2307815

2. Nora Broy, Mengbing Guo, Stefan Schneegass, Bastian Pfleging, and Florian Alt. 2015. Introducing Novel Technologies in the Car: Conducting a Real-world Study to Test 3D Dashboards. In *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. ACM, New York, NY, USA, 179–186. DOI: http://dx.doi.org/10.1145/2799250.2799280

3. Florian Echtler and Maximilian Häußler. 2018. Open Source, Open Science, and the Replication Crisis in HCI. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems (CHI EA '18)*. ACM, New York, NY, USA. DOI: http://dx.doi.org/10.1145/3170427.3188395

4. Niels Henze, Enrico Rukzio, and Susanne Boll. 2011. 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)*. ACM, New York, NY, USA, 133–142. DOI: http://dx.doi.org/10.1145/2037373.2037395

5. Martin Schmettow. 2012. Sample Size in Usability Studies. *Commun. ACM* 55, 4 (April 2012), 64–70. DOI: http://dx.doi.org/10.1145/2133806.2133824