The Renaissance of Wizard of Oz (WoOz) – Using the WoOz methodology to prototype automated vehicles

Klaus Bengler[1], Kamil Omozik[2], & Andrea Isabell Müller[1]
[1]Technical University of Munich, Chair of Ergonomics, [2]BMW Group
Germany

Abstract

The strong increase in momentum behind the development of automated systems is leading to a change in paradigm with regard to the distribution of control in human-machine interaction. Therefore, in the context of automated driving, it is necessary to explore fundamental questions such as the interaction between driver and vehicle. However, the underlying automated driving functions are still under development and thus can only be used for studies to a limited extend. From a technical point of view, the introduction of automated systems results in an increased proportion of probabilistic components. Due to the resulting non-determined behaviour of the automation, it is difficult to perform studies in a systematic manner. A suitable method to study the effects of such "intelligent" probabilistic systems are Wizard of Oz (WoOz) setups, where a human simulates the behaviour of the system. The results obtained through WoOz studies are promising, but considering the system behaviour reproduced by the driving wizard researchers apply the method in different ways. Furthermore, there seems to be a lack of systematics regarding the experimental procedure, ethics and the guarantee of scientific quality. This article evaluates and systematizes published experimental approaches and proposes a specification language for the driving wizard's behaviour.

Introduction

The introduction of automated vehicles is leading to fundamental changes in the relation between vehicles, users and other traffic participants. To analyse this change in relation real automated vehicles can only be used to a limited extend, since the underlying driving functions are still under development. At the same time, developers of the technical system need input on human abilities and restrictions in interaction, which cannot simply be transferred from other domains like aviation or process control. Gasser et al. (2015) give a detailed overview of relevant questions related to level 3 automated driving. Additionally, the rise in automated driving functions within the vehicle system leads to an increased proportion of probabilistic components. However, from the perspective of human factors research, a more deterministic behaviour of the technical system, i.e. the automation, is necessary to evaluate human-machine interaction, since investigations could suffer from random effects in scene interpretation, environmental influences or surrounding traffic behaviour.

In D. de Waard, A. Toffetti, L. Pietrantoni, T. Franke, J-F. Petiot, C. Dumas, A. Botzer, L. Onnasch, I. Milleville, and F. Mars (2020). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2019 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

A comparable situation was given in the area of human-computer interaction when, for example, intelligent tutoring systems and speech or gesture recognition were mature to be introduced but had to be evaluated in a systematic way. Here, the Wizard of Oz (WoOz) paradigm was applied with great success and enabled research on human-machine interaction in parallel to technical development. Within the automotive research community, WoOz vehicles are also an established method for analysing the effects of "intelligent" probabilistic systems that have not been fully developed yet, such as automated vehicles.

Exemplary application of WoOz studies

WoOz studies are used when complex systems have to be evaluated prior to becoming available. The systems are simulated by humans, the so-called wizards (Fraser & Gilbert, 1991), in a hidden manner. Ideally, this causes users to believe that they are interacting with the real technical systems rather than a simulated one (Bernsen et al., 1994). John F. Kelley invented the WoOz paradigm in 1975 to simulate a not yet functional speech recognition system (Green & Wei-Haas, 1985). Further studies have followed using the WoOz paradigm to simulate natural language recognition systems, such as Kelley (1983) simulating a software assistant to support users when interacting with a digital calendar programme or Gould et al. (1983) simulating a "listening typewriter". From the early 1990s on the WoOz paradigm was also used to prototype multi-modal recognition systems. Hauptmann (1989) simulated a graphics programme that could be used to edit images through speech and gestural input, while Robbe et al. (1997) simulated a spatial planning programme that could likewise be controlled through speech and gestural input.

In the automotive sector, the WoOz methodology is commonly used to design user interfaces (Pettersson & Ju, 2017), such as a multi-modal recognition system to control non-driving related vehicle functions (Stecher et al., 2018). However, the WoOz methodology can also be used to simulate automated vehicles. In this case, so-called driving wizards (Baltodano et al., 2015), simulate the automation by driving the vehicle hidden from participants (Coelingh et al. 2018). When simulating natural language or multi-modal recognition systems, the wizards do not sit in the same room as the participants and the system to be simulated (Hauptmann, 1989; Stecher et al., 2018). However, when simulating automated vehicles, driving wizards act as part of the test tool and are located within the test tool (Müller et al., 2019) allowing them to experience their actions in the same way as the participants.

In 2006 Kiss et al. (2006) developed a WoOz vehicle for the first time to simulate driver assistance systems in real traffic conditions. In the same year, Schomerus et al. (2006) developed the theatre-system technique, which is set in a driving simulator and represents a special case: the deception used in WoOz studies can deliberately be lifted so that researchers can directly get in touch with participants (Schomerus et al. 2006). Fuelled by the development of the Ghost Driver methodology (Rothenbücher et al. 2015) and the RRADS vehicle setup (Baltodano et al., 2015), the WoOz paradigm is currently becoming more used to simulate automated vehicles in real traffic.

Common construction forms of WoOz vehicles

Studies involving vehicle occupants as participants require complex vehicle setups to create the illusion of an automated vehicle. All these setups have in common that usually a participant, a driving wizard and an interaction wizard occupy the vehicle. The interaction wizard typically also acts as the investigator. The classification by Manstetten et al. (2019) does not cover all published WoOz vehicle setups. Therefore, a more systematized approach is proposed in the following.

WoOz vehicle setups used for occupant studies can be divided into setups where the participant is seated in the front row or in the back row. Vehicle setups, where the participant is seated in the back (see Figure 1), are typically used for simulating level 5 automation (Karjanto et al., 2018; Sandhaus & Hornecker, 2018; Sherry et al., 2018). The driving wizard operates the vehicle by using the serial driver workplace. An opaque partition obscures the vehicle controls and the driving wizard. Visibility to the front of the vehicle for participants can be realised by mounting a TV displaying a video of the environment (Karjanto et al., 2018) or by not covering the area between the headrests and the vehicle roof (Sherry et al., 2018).

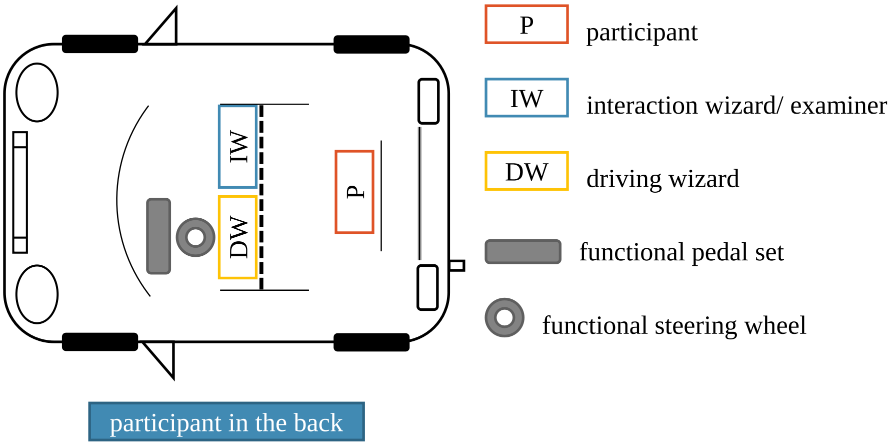

Figure 1: Common WoOz vehicle setup where participants are seated in the back row

In case of participants sitting in the front row four different WoOz vehicle setups could be identified (see Figure 2). These can be divided into setups, where participants can drive the vehicle (Figure 2 bottom row) and ones where they cannot (Figure 2 top row).

Setups where participants cannot drive the vehicle should be used for studying level 4 or level 5 automation since Requests to Intervene (RtI) typically cannot be represented. One of these setups is based on a left-hand drive vehicle (see Figure 2 top left). To ensure the disbelief, the driving wizard and vehicle controls are hidden using a partition between driving wizard and participant.

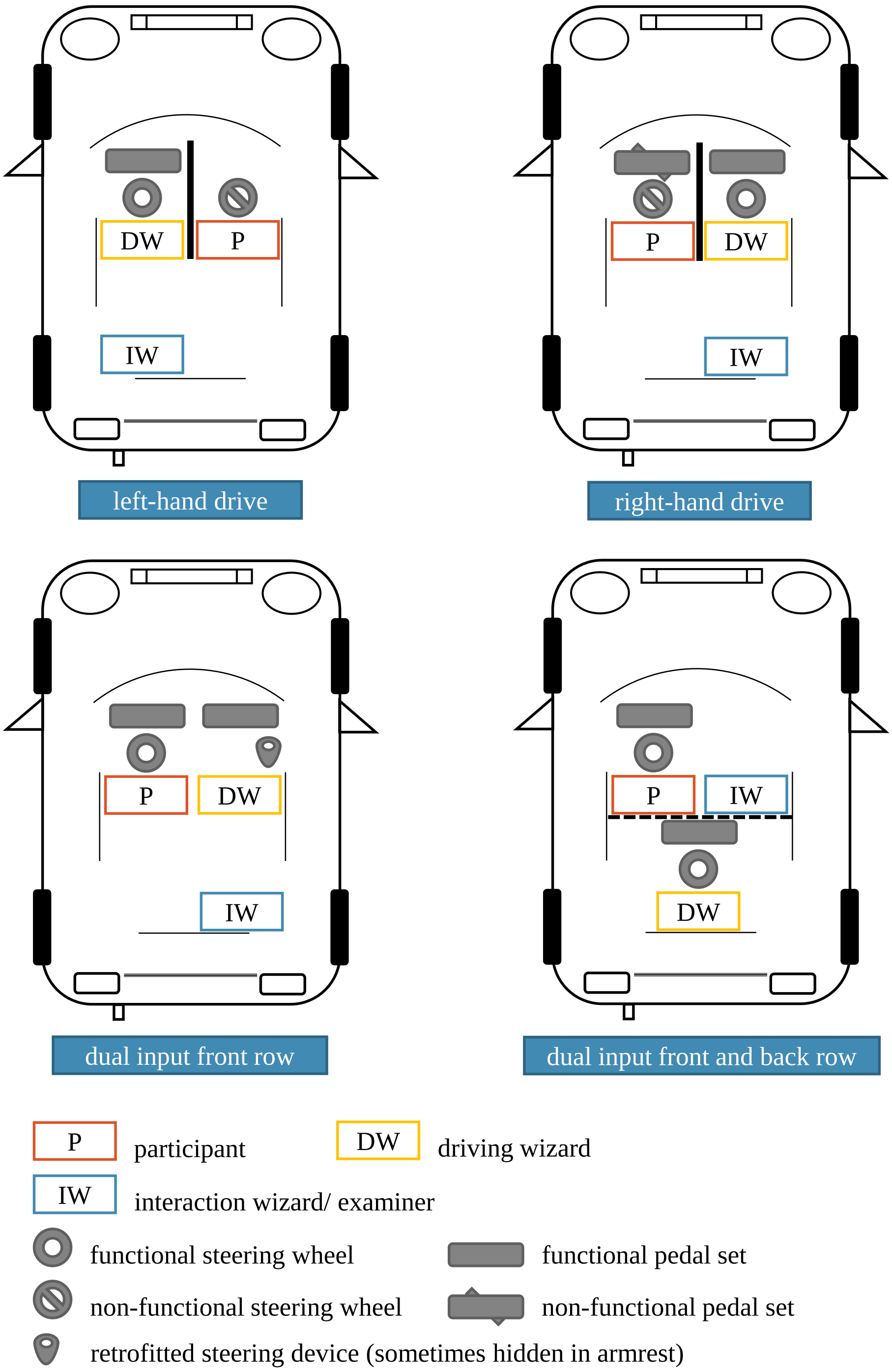

Figure 2. Common WoOz vehicle setups where participants are seated in the front row

Furthermore, the participant's seat is equipped with a non-functional steering wheel. Baltodano et al. (2015) developed this setup called RRADS (Real Road Autonomous Driving Simulator). Another setup is based on a right-hand drive vehicle (see Figure 2 top right). The seating position of participants on the (in most countries) usual driver's side acts as a strong cue, that participants are not only passengers. To intensify this feeling, the driving wizard and vehicle controls are concealed using either a partition (Wang et al., 2017), a curtain (Weinbeer et al., 2017) or a hat with covers on the right side (Rittger et al., 2017). To simulate a level 3 automation, Wang et al. (2017) invented the Marionette system, where the driving wizard reproduces the exact input that participants perform using dummy control elements. Weinbeer et al. (2017) attached three displays to the dashboard that represented the highway lanes to simulate an RtI to which participants had to react using dummy control elements.

In vehicle setups where participants are capable of driving themselves, they are always provided with the serial vehicle control elements, whereas the driving wizard uses a retrofitted driving environment. The driving wizard can sit either in the front row (see Figure 2 bottom left) or in the backrow (see Figure 2 bottom right). These setups can be used to simulate automation levels 2 to 4. Level 5 can be simulated with certain limitations since the serial driving workplace provides a strong cue of needing to control the vehicle at some point.

A dual front row input can be realised by providing the driving wizard with another set of pedals and a hidden steering device integrated into the right door (Naujoks et al., 2019). In this case, there is no visibility barrier to ensure that the vehicle is always either controlled by the driving wizard or the participant during simulated RtIs. When using a retrofitted steering wheel as a steering option for driving wizards, a visibility barrier is installed to improve the illusion of an automated vehicle. However, to ensure safe transfers of control during RtIs, the driving wizard must be provided with a display of the current state of vehicle control (Sportillo et al., 2019). For WoOz vehicles where the driving wizard is seated in the back, a semi-transparent glass, that allows the driving wizard to view through the windscreen, separates the driving wizard and the participant (Jarosch et al., 2019). As a special feature of this vehicle setup, participants can sit completely by themselves in the front of the vehicle (Osz et al., 2018).

Simulating automated driving behaviour

To simulate automated driving behaviour, the driving wizard must be able to consistently reproduce an automated driving style. For this reason, it is advisable to define the automated driving style and instruct driving wizards accordingly. The most obvious instruction for driving wizards is to let them drive similar to their idea of how automated vehicles will behave (Wang et al., 2017). Moreover, it is possible to instruct driving wizards in a metaphoric way, e.g. by telling them to drive "similar to a professional limo driver" or to achieve a smooth and conservative driving style (Baltodano et al., 2015). Additionally, a qualitative description of the intended driving style can be used. Possible parameters to be defined include the accelerating and decelerating behaviour (Baltodano et al., 2015), the stopping behaviour (Ekman et al., 2019), the distance to surrounding traffic (Ekman et al., 2019), the choice of lane (Naujoks et al., 2019), the lane change behaviour (Weinbeer et al., 2018), the choice of gear (Ekman et al., 2019) as well as the position within a lane (Ekman et al., 2019). The most detailed way of instructing driving wizards is to specify driving strategies of automated vehicles by quantitative parameter sets. These can refer to the maximum velocity (Jarosch et al., 2019; Naujoks et al., 2019; Omozik et al., 2019; Weinbeer et al., 2018), a maximum lateral acceleration (Karjanto et al., 2018) or permitted ranges for longitudinal acceleration and deceleration (Ekman et al., 2019). To realize the predefined driving behaviour, Adaptive Cruise Control (ACC) and Lane Keeping Assistant can be used (Rittger et al., 2017). However, one must be aware that this holds the risk of unintentionally simulating a state-of-the-art system (Weinbeer et al., 2018).

Methodology and good practices

Through the presence of a human wizard, a variety of cognitively demanding tasks, that have not been implemented yet, can be simulated (Bernsen et al., 1994) to realise novel systems fast and without technical development (Kiss et al., 2006). The WoOz methodology allows for a timely user feedback as well as observations in a natural environment and is cost-efficient (Stevens et al., 2019). Compared with existing automated systems, the WoOz paradigm allows for less constrained experiments by using improvisation through the wizard, but also more systematically constrained experiments by omitting the limitations of an automated system (Osz et al., 2018). Since the later technical realisation of the system is unclear (Stevens et al., 2019), one methodological risk is to simulate the technical system in an idealistic way or to insert human deficits into the simulation of machine-like behaviour. Furthermore, the wizard is in a feedback loop with the surrounding traffic system. Compared to other WoOz realizations this is a novelty. In general, it seems challenging to ensure the scientific quality of results achieved using the WoOz paradigm. In this context, Müller et al. (2019) identified the following main methodological challenges related to WoOz:

1. Participants must be under the impression that they are interacting with a real automated vehicle.
2. The simulated automated vehicle must behave as if it were a real automated vehicle.
3. One driving wizard must be able to reproduce the pre-defined driving style at different times.
4. Different driving wizards must be able to reproduce the pre-defined driving style.

As a result, when using the WoOz methodology, not only hypotheses considering the research questions have to be tested, but also considering the comparability of test drives and the believability of the illusion. Therefore, to ensure reliable and comparable results, WoOz requires investigators to record, analyse and report additional data compared to other research paradigms (Dillmann et al., 2019). These include the driving dynamics produced by the driving wizard, the speed and location of surrounding traffic participants, as well as the interface output displayed to participants. Interviews and video recordings are useful to evaluate how participants experienced the illusion created through WoOz (Maulsby et al. 1993). Moreover, not all kinds of research questions can or should be answered by employing the WoOz methodology. It should not be used for research questions where an input by the driver triggers a system reaction and when an introduction of a specific system in the market has to be decided. In addition, take-over situations always have to be manageable and therefore cannot be examined in situations with high urgency (Feldhütter et al., 2017).

Need for research

The authors propose an “inverted” Turing Test methodology to be able to validate different driving wizards related to the research question under investigation and in relation to the automation system under development. Furthermore, a taxonomy is needed to describe the wizard’s driving and decision behaviour in a qualitative and quantitative way. As it seems challenging to instruct and for the driving wizard to monitor quantitative values while driving, a qualitative description seems reasonable to instruct wizards in a metaphoric way on driving style and strategic behaviour on the manoeuvring level. A quantitative description of the wizard’s driving behaviour is necessary to enable a quantitative comparison between different data sets of one or more wizards and with the automated system under investigation. It is informative to compare different data sets using average values. However, a more differentiated view on values gathered before, during and after certain manoeuvres seems to be necessary. For this comparison, several metrics seem suitable, such as the minimum time to collision (TTC_min), the frequency of lane changes, the minimum gap size of a lane change as well as metrics quantifying the cooperation with other road users. Besides objective data regarding the manoeuvres, it also seems necessary to describe the environment at the time of the manoeuvre. This could be traffic density, number of road lanes and time of day as well as weather and road conditions. Currently, there is no criteria available to decide systematically between data sets produced by different driving wizards in similar contexts or by the same driving wizard in differing contexts. This problem is well known from field operational tests and naturalistic driving studies. Systematic comparison and selection criteria should be checked for a potential transfer. Additionally, it seems necessary to compare the simulated driving behaviour created by driving wizards with that of a real automated vehicle.

Conclusion

The WoOz paradigm was invented in 1975 to prototype natural language recognition systems. Nowadays it is becoming increasingly more popular to simulate automated vehicles on real roads. Typical WoOz vehicle setups were identified, including a setup where participants are seated in the back, setups based on left-hand drive as well as right-hand drive vehicles and two setups where both the participant and the driving wizard can drive the vehicle. The identified strategies to instruct driving wizards can be divided into metaphoric, qualitative and quantitative instructions. Lastly, strengths and weaknesses of the WoOz paradigm were discussed, possible fields of application were evaluated and a further need for research to improve the scientific quality of WoOz studies was determined.

References

Baltodano, S., Sibi, S., Martelaro, N., Gowda, N., & Ju, W. (2015). The RRADS Platform: A Real Road Automated Driving Simulator. In *Proceedings of the*

7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 281-288). New York: ACM.

Bernsen, N.O., Dybkjær, H., & Dybkjær, L. (1994). Wizard of Oz Prototyping: When and How? In *CCI Working Papers in Cognitive Science and HCI*.

Coelingh, E., Nilsson, J., & Buffum, J. (2018). Driving tests for self-driving cars. *IEEE Spectrum*, *55*, 40-45.

Dillmann, J., den Hartigh, R., Kurpiers, C., Raisch, F., de Waard, D., & Cox, R. (2019). Intentional Dynamics in Conditionally Automated Driving. In *International Conference On Perception And Action*.

Ekman, F., Johansson, M., Bligård, L.-O., Karlsson, M., & Strömberg, H. (2019). Exploring automated vehicle driving styles as a source of trust information. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 268-279.

Feldhütter, A., Hecht, T., & Bengler, K. (2017). *Fahrerspezifische Aspekte beim hochautomatisierten Fahren* (Report FE 82.0628/2015). Munich, Germany: Technical University of Munich, Chair of Ergonomics.

Fraser, N.M., & Gilbert, G.N. (1991). Simulating speech systems. *Computer Speech and Language*, *5*, 81-99.

Gasser, T.M., Schmidt, E.A., Bengler, K., Chiellino, U., Diederichs, F., Eckstein, L., Flemisch, F., Fraedrich, E., Fuchs, E., Gustke, M., Hoyer, R., Hüttinger, M., Jipp, M., Köster, F., Kühn, M., Lenz, B., Lotz-Keens, C., Maurer, M., Meurer, M., Meuresch, S., Müller, N., Reitter, C., Reschka, A., Riegelhuth, G., Ritter, J., Siedersberger, K.-H., Stankowitz, W., Trimpop, R., & Zeeb, E. (2015). Report on the Need for Research: Round Table on Automated Driving - Research Working Group.

Gould, J.D., Conti, J., & Hovanyecz, T. (1983). Composing letters with a simulated listening typewriter. *Communications of the ACM*, *26*(4), 295-308.

Green, P., & Wei-Haas, L. (1985). *The Wizard of Oz: A Tool for Rapid development of User Interfaces* (Report UMTRI-85-27). Ann Arbor, USA: University of Michigan, Transportation Research Institute.

Hauptmann, A.G. (1989). Speech and Gesture for Graphic Image Manipulation. In *Proceedings of the ACM CHI'89 Human Factors in Computing Systems Conference* (pp. 241-245). New York, ACM.

Jarosch, O., Paradies, S., Feiner, D., & Bengler, K. (2019). Effects of non-driving related tasks in prolonged conditional automated driving – A Wizard of Oz on-road approach in real traffic environment. *Transportation Research Part F: Traffic Psychology and Behaviour*, 65, 292-305.

Karjanto, J., Yusof, N. M., Wang, C., Terken, J., Delbressine, F., & Rauterberg, M. (2018). The effect of peripheral visual feedforward system in enhancing situation awareness and mitigating motion sickness in fully automated driving. *Transportation Research Part F: Traffic Psychology and Behaviour*, 58, 678-692.

Kelley, J.F. (1983). An empirical methodology for Writing User-Friendly Natural Language computer applications. In Janda, A (Ed.), *CHI '83 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 193-196). New York, ACM.

Kiss, M., Schmidt, G., & Babbel, E. (2006). *Das Wizard of Oz Fahrzeug: Rapid Prototyping und Usability Testing von zukünftigen Fahrerassistenzsystemen.* VW Konzernforschung.

Manstetten, D., Marberger, C., & Beruscha, F. (2019). Wizard-of-Oz Experiments in Real Traffic - Can They Restart Human Factors? In *9. Darmstädter Kolloquium „mensch + fahrzeug"* (pp. 21-31).

Maulsby, D., Greenberg, S., & Mander, R. (1993). Prototyping an Intelligent Agent Through Wizard of Oz. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 277-284). New York: ACM.

Müller, A.I., Weinbeer, V., & Bengler, K. (2019). Using the Wizard of Oz Paradigm to Prototype Automated Vehicles: Methodological Challenges. In *Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings* (pp. 181-186). New York: ACM.

Naujoks, F., Purucker, C., Wiedemann, K., & Marberger, C. (2019). Noncritical State Transitions During Conditionally Automated Driving on German Freeways: Effects of Non–Driving Related Tasks on Takeover Time and Takeover Quality. *Human Factors,* 61, 596-613.

Omozik, K., Yang, Y., Kuntermann, I., Hergeth, S., & Bengler, K. (2019). How long does it take to relax? Observation of driver behaviour during real-world conditionally automated driving. In *Proceedings of the Tenth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 245-251).

Osz, K., Rydström, A., Fors, V., Pink, S., & Broström, R. (2018). Building Collaborative Test Practices: Design Ethnography and WOz in Autonomous Driving Research. *Interaction Design and Architecture(s) Journal*, 12-20.

Pettersson, I., & Ju, W. (2017). Design Techniques for Exploring Automotive Interaction in the Drive towards Automation. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (pp. 147-260). New York: ACM.

Rittger, L., Wiedemann, K., Schmidt, G., & Green, C.A. (2017). HMI for anticipation of upcoming curvature in automated lateral control. In M. Lienkamp (Ed.), *8. Tagung Fahrerassistenz, Einführung hochautomatisiertes Fahren*.

Robbe, S., Carbonell, N., & Dauchy, P. (1997). Constrained vs spontaneous speech and gestures for interacting with computers: A comparative empirical study. In S. Howard, J. Hammond, & G. Lindgaard (Eds.), *Human-Computer Interaction. IFIP — The International Federation for Information Processing* (pp. 445–452). Boston: Springer.

Rothenbücher, D., Mok, B., Li, J., Ju, W., & Sirkin, D. (2015). Ghost Driver: A Platform for Investigating Interactions Between Pedestrians and Driverless Vehicles. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 44-49). New York: ACM.

Sandhaus, H., & Hornecker, E. (2018). A WOZ Study of Feedforward Information on an Ambient Display in Automated Cars. In *The 31st Annual ACM Symposium on User Interface Software and Technology Adjunct Proceedings* (pp. 90-92). New York: ACM.

Schomerus, J., Flemisch, F., Kelsch, J., Schieben, A., & Schmuntzsch, U. (2006). Erwartungsbasierte Gestaltung mit der Theatersystem-/ Wizard-Of-Oz-Technik am Beispiel eines haptischen Assistenzsystems. In *AAET 2006 Automatisierungssysteme, Assistenzsysteme und eingebettete Systeme für Transportmittel* (pp. 209-225).

Sherry, J., Beckwith, R., Esme, A.A., & Tanriover, C. (2018). Getting things done in an automated vehicle. In *Social Robots in the Wild Workshop at the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction*.

Sportillo, D., Paljic, A., & Ojeda, L. (2019). On-Road Evaluation of Automated Driving Training. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 182-190). Piscataway: IEEE.

Stecher, M., Michel, B., & Zimmermann, A. (2018). The Benefit of Touchless Gesture Control: An Empirical Evaluation of Commercial Vehicle-Related Use Cases. In N.A. Stanton (Ed.), *Advances in Human Aspects of Transportation. AHFE 2017. Advances in Intelligent Systems and Computing* (pp. 383–394). Cham: Springer.

Stevens, G., Meurer, J., Pakusch, C., & Bossauer, P. (2019). Investigating Car Futures from Different Angles: An Overview of Methods Used to Study Human Factors of Autonomous Driving. In *Mensch und Computer 2019 – Workshopband* (pp. 400–409). Bonn: Gesellschaft für Informatik e.V.

Wang, P., Sibi, S., Mok, B., & Ju, W. (2017). Marionette: Enabling On-Road Wizard-of-Oz Automated Driving Studies. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 234-243). New York: ACM.

Weinbeer, V., Baur, C., Radlmayr, J., Bill, J.-S., Muhr, T., & Bengler, K. (2017). Highly automated driving: How to get the driver drowsy and how does drowsiness influence various take-over aspects? In M. Lienkamp (Ed.), *8. Tagung Fahrerassistenz, Einführung hochautomatisiertes Fahren*.

Weinbeer, V., Muhr, T., & Bengler, K. (2018). Automated Driving: The Potential of Non-driving Related Tasks to Manage Driver Drowsiness. In S. Bagnara, R. Tartaglia, S. Albolino, T. Alexander, & Y. Fujita (Eds.), *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018): Volume VI: Transport Ergonomics and Human Factors (TEHF), Aerospace Human Factors and Ergonomics* (pp. 179-188). Cham: Springer.