Towards a Cognitive Model of the Takeover in Highly Automated Driving for the Improvement of Human Machine Interaction.

Marlene Scharfe (m.scharfe@campus.tu-berlin.de)
Department of Psychology and Ergonomics, Marchstr. 23, 10587 Berlin, Germany

Nele Russwinkel (nele.russwinkel@tu-berlin.de)
Department of Psychology and Ergonomics, Marchstr. 23, 10587 Berlin, Germany

Abstract

In this study, an ACT-R cognitive model, that depicts cognitive dynamics during a takeover in highly automated driving is developed. Such a model is inevitable to explain strong differences between studies and display and represent cognitive dynamics. Thus, the goal of the model is, to validly display the time sequence of the steps, that are undertaken to build up situation awareness during the takeover task. It is developed based on video analysis and pertinent literature. Empirical data of a real traffic study show takeover patterns that are applied after engaging into a non-driving related task. Correlations between model predictions and empirical data evaluation show, that the model is able to display cognitive dynamics. It serves as solid basis, but still further development is aspired, concentrating on the impact of traffic complexity.

Keywords: highly automated driving; HAD; cognitive modeling; ACT-R; takeover; TOR; conditional automation; NDRT; non-driving related tasks; real vehicle study; resource model; situation awareness; SA;

Introduction

The development of technological innovations in the field of highly automated driving is growing rapidly. The next level of automation (SAE Level 3; SAE, 2014) enables the driver to engage into non-driving related tasks (NDRT) during the automated drive. Still, the driver needs to respond properly to a takeover request (TOR; SAE, 2014). Thus, the driver has to be enabled to take over the driving task in a safe and comfortable manner. Several studies have investigated takeover times (Feldhütter, Gold, Schneider, & Bengler, 2017; Gold, Damböck, Lorenz, & Bengler, 2013; Naujoks & Neukum, 2014; Walch, Lange, Baumann, & Weber, 2015). As various factors influence the takeover, current results concerning takeover times and behavior in Level 3 illustrate incomparable results that range from 1,14s (Zeeb, Buchner, & Schrauf, 2015) to 15s (Merat, Jamson, Lai, Daly, & Carsten, 2014). It is of scientific relevance to understand underlying cognitive dynamics that lead to these differences. Also, to improve the human-machine interaction (HMI) and the product development in the indusrty, it is inevitable to unravel the black box of cognitive dynamics and gain an understanding of how the human processes the takeover and builds up situation awareness (Endsley, 1995). According to Endsley (1995), situation awareness encompasses three main stages: *perception*, *comprehension* and *projection* (included in Figure 1). The basis is formed by a persons visual perception of the environment. Based on the perception, the meaning of the current situation has to be understood (comprehension) and a future status is projected (Endsley, 1995). The aim of this study is to develop a cognitive model, representing cognitive dynamics during the takeover task. Yet, cognitive models for the driving task exist (e.g. Salvucci, 2006), but no renowned cognitive model explicitly displays the takeover and includes surrounding traffic.

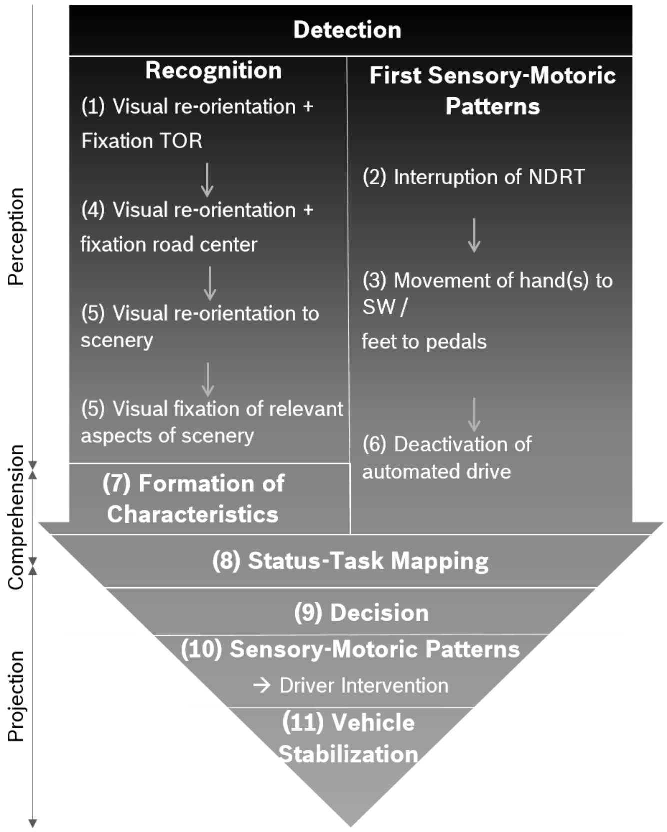

Figure 1: Definition of the Takeover Process and the Phases of Situation Awareness for the cognitive model approach (Source: own figure).

In this paper a first cognitive modeling approach of the takeover task is introduced and a comparison of model results to empirical data is drawn. The goal is to create an understanding of cognitive mechanisms during the separate steps of the takeover process (visualized in Figure 1). Cognitive processes are later described in detail. The focus of the model lies on the perception mechanisms rather than the interruption process which has been modeled by Borst, Taatgen, and van Rijn (2015). To better understand perception mechanisms, the interruption is implemented here simply as a new goal setting, which can later be expanded. Here, the steps to update situation awareness and perform the takeover task are modeled. As soon, as the the basic cognitive dynamics are understood, aspects of complexity will be included in future work.

For the current approach, it is investigated whether a cognitive model is able to a.) explain cognitive mechanisms during the takeover according to empirical data, b.) depict mechanisms that arise due to NDRTs and c.) illustrate cognitive processes to update SA using the visual component.

Methods

The approach of this study is to gain an understanding of cognitive dynamics during the takeover task. Thereby, behavior during a takeover can be predicted and empirical results of past studies can be explained. Based on pertinent concepts and results of relevant studies, the takeover process in conjunction with the three stages of situation awareness (SA; Endsley, 1995) is defined (Figure 1) and a cognitive model established for the takeover task. In order to validate overall predictions of the model, these are compared to empirical data of a driving study in a real traffic environment. After validating the overall model performance, different conditions of the NDRT are closer examined to extract potential improvements. Following, the data acquisition is described and methods as well as the functions of the model outlined.

Data Acquisition

The utilized data derives from a real traffic study of another project (Ko-HAF, 2017) in 2017 with a Wizard of Oz vehicle in the area of Stuttgart, Germany. The vehicle allows the experience of highly automated driving in a real driving environment. The co-driver is able to drive the car covertly via a control at the passengers seat and thus simulate highly automated driving (Level 3). This results in the drivers ability to turn away from the driving task (Ko-HAF, 2017) and engage into a non-driving related task (NDRT). For the current approach, the type of NDRT that is performed, is not relevant. As soon as a TOR is triggered, participants have to take over. For this study's purpose $N = 28$ participants are evaluated. Half of the subjects ($N = 14$) are used for qualitative video analysis concerning gaze- and motoric behavior. This data is used for the model development. Two raters independently view the videos and assess the participants' behavior in a standardized way. Four behavioral relevant steps are extracted (1. Gaze TOR, 2. Interruption NDRT, 3. Hands steering wheel (SW) and 4. Gaze Road), that define the time course of the takeover for the current purpose. In order to validate resulting model predictions, the other half ($N = 14$) is quantitatively evaluated concerning the four steps and compared to model results.

Cognitive Model

The cognitive model displays sub-steps that are undertaken during the takeover and cognitive dynamics for visual perception, motoric reactions and decision making. The aim is, to have model predictions in accordance with empirically found reaction times. Motoric, visual and cognitive steps to update SA are included into the model (Figure 1). To realize the implementation, the cognitive architecture ACT-R (Anderson et al., 2004) is used. It allows the modeling of a wide range of higher cognitive processes (Taatgen, 1999) and provides an accurate representation of human abilities (Salvucci, Boer, & Liu, 2001). ACT-R contains various modules (intentional, declarative, visual and manual) that communicate with each other through buffers (Anderson et al., 2004). The opportunity to model processes of the particular modules separately and gain insights of their interaction in certain situations is essential for modeling the takeover. Cognitive modules form the mental representation elaboration and decision-making processes on attentional and automatic levels (Bellet et al., 2012). Visual and aural modules describe the visual and aural processing of stimuli (Anderson et al., 2004). The manual module is responsible for the execution of motoric responses.

For this study's' purpose, three steps are undertaken. Based on literature (Endsley, 1995; Salvucci, 2006) and on qualitative analysis of video material from the driving study ($N = 14$), first the process that occurs during a takeover is defined for the current model (Figure 1). These steps are then used for the definition of the cognitive model. Third, the model is tested and compared to empirical data. It is important to clearly define the takeover task and steps that are undertaken to regain control. (Figure 1) outlines how this is addressed. Additionally, it has to be understood how the different stages of SA (Endsley, 1995) are realized to implement corresponding cognitive structures in ACT-R. The focus of the current model lies on the visual *perception phase* of SA as it is the most important of the three SA stages in driving, the first to occur (Ratwani, McCurry, & Trafton, 2010) and forms the basis for the subsequent steps. Following, the takeover process and the corresponding realization in the model are described.

The scenario, that the model illustrates, describes a highly automated driving situation (Level 3) on a three lane highway. The model starts with a NDRT during the automated drive and ends with an action decision after the takeover. Figure 2 illustrates the model productions.

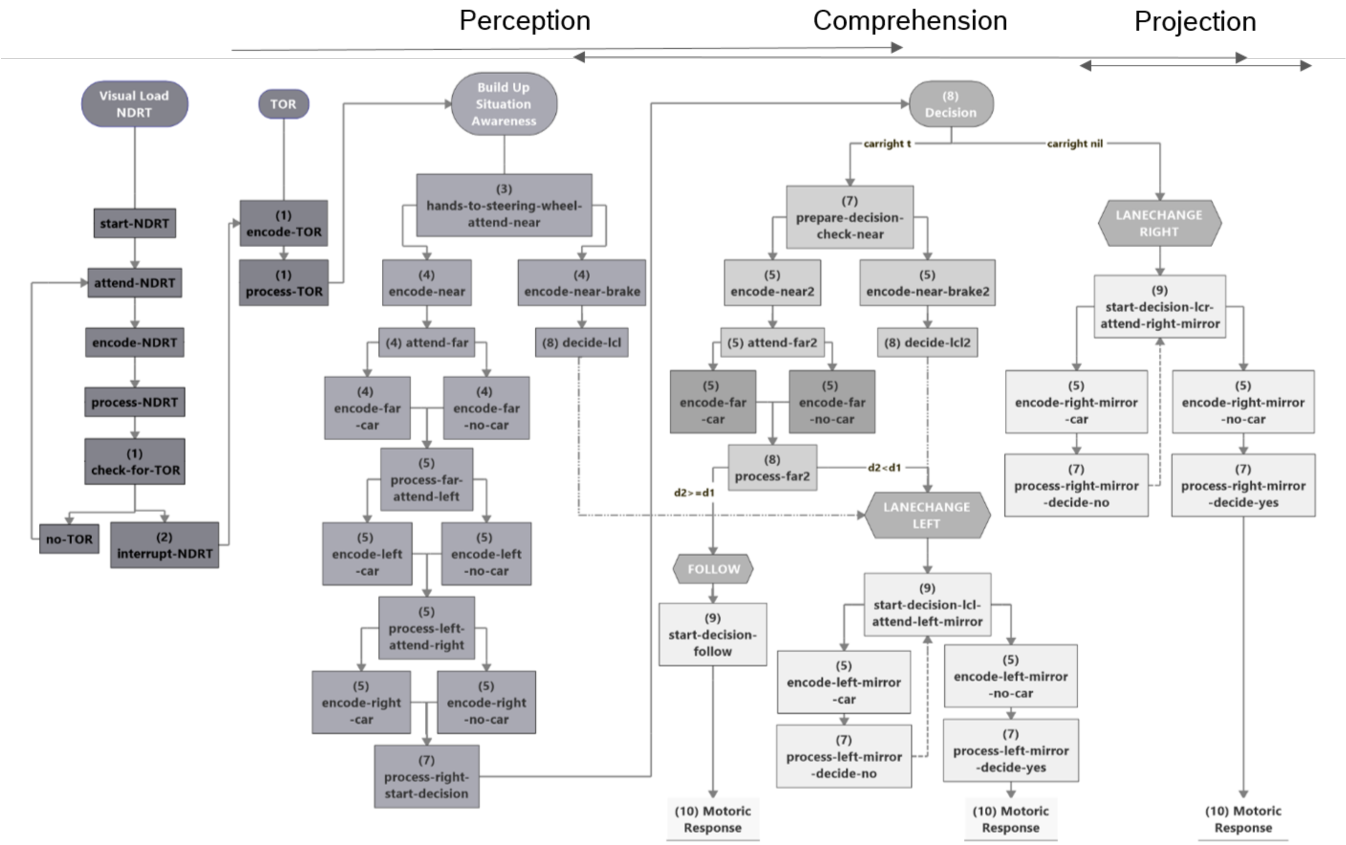

Figure 2: Cognitive Model of the Transition Process after an Automated Drive. Representation of productions from the NDRT to the motoric response (Source: own figure).

Figure 3 represents the environment, the model interacts with and the areas that are attended by the model. The visual perception undergoes three main steps in the *perception phase* (Endsley, 1995). While focusing on the NDRT, the model is constantly checking for a takeover request (TOR) in the visual-location. This is due to the given task of taking over the driving as soon as a TOR appears. Due to the perception of a visual or aural stimulus, the gaze is moved to the TOR (1.; Visual reorientation and fixation of the TOR) and the goal set to attend the TOR. In case no TOR can be detected, the NDRT is started again. Otherwise as soon as the TOR is triggered, the NDRT (Visual Load NDRT) has to be interrupted and a retrieval request for the meaning of the TOR is made. At the same time, the visual interruption of the NDRT (2.) is executed. It automatically occurs with the attentional shift to the TOR in the *perception phase* (Endsley, 1995; Salvucci, 2006). They come along with first automated sensory-motoric steps, encompassing the movement of hands to the steering wheel and feet to the pedals (3.). In the model only the movement of the hands to the steering wheel can be implemented, as ACT-R does not include the feet yet. Nevertheless, a movement action is carried out by the manual module. These are based on an automatic reaction rather than focused decision making and fluently merge into the *comprehension phase* (Endsley, 1995).

In separate productions, the TOR is attended, encoded and processed. After the TOR has been processed, SA has to be build up (Build Up Situation Awareness). Hence, the gaze is oriented to the road center and the front lane (near and far area) is checked for objects (4.; Salvucci, 2006). In case an object is detected in the near area (encode-near-brake), a strong brake is carried out (motoric module) and a direct decision to change to the left lane is made, setting the goal to lanechange left. This case pictures a critical event and should not apply when a non-critical TOR is triggered. Thus, in the current context this case is postponed, as the study focuses on non-critical takeover scenarios. Nevertheless, for completeness of the model it is still necessary to cover the scenario. In the majority of cases though, no object is in the near area and the far front lane is attended. It is encoded whether there is a car or not (attend-far) and the result processed. After this, the rest of the scenery is attended (5.), starting with the left lane (process-far-

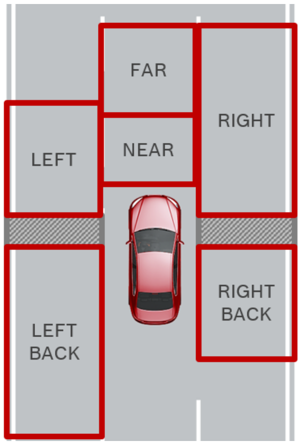

Figure 3: Focus Areas of the Cognitive Model for Action Decision (Source: own figure).

attend-left). Again, the model encodes whether there is a car or not. The result is processed and stored in a chunk in the imaginal. This process is repeated to perceive the status of the right lane likewise (process-left-attend-right). Each status (object or not) of the attended areas is stored in the imaginal to form a representation of the traffic environment. Here, characteristics of the situation are formed and the current status is recognized (7.; Endsley, 1995) while the automation is deactivated (6.). The deactivation of the automation is not represented in the model, as there is no common function yet and operates by oversteering. After a mapping between the perceived status and the task (8.), an action decision is made (9.; *comprehension phase* of SA; Endsley, 1995). The decision productions (Decision) are based on the status of the surrounding traffic in the focus areas (Figure 3) and the underlying law. In case no car has been detected on the right lane, a lane change to the right is triggered (LANECHANGE RIGHT). Thus, the right mirror is attended by the visual module to check, whether the right back is free. If it is free, the lane change is executed by a motoric response of the manual module (10.). In case, there is a car in the right back, the mirror is attended repeatedly, until the lane change can be performed safely. If the right lane is not free, the front road is attended again, applying the same mechanisms as before. Now, distances from the first allocation that are stored in the imaginal are compared to the current perception (process-far2). If the distance to the car in the front stays the same or is increasing, car following is decided and performed (FOLLOW). If the distance to the front car is decreasing, a lane change to the left is triggered (LANECHANGE LEFT). Again, the mirror is repeatedly checked for occupation of the left back. Equally to the right lane change, a left lane change is only performed by the manual module, when it can be carried out safely (no object in the visual-location for the left back; Figure 2). The model runs until this stage, predicting type and time of action execution in different traffic conditions. Still, the focus of this paper is to validate the model steps in order to produce valid predictions. After the driver intervention (10.), the vehicle is stabilized (11.; *projection phase* of SA; Endsley, 1995).

Results

For the statistical evaluation, only data that the model does not base on, is used. Four timestamps were defined for the individual takeover behavior (1. Gaze TOR, 2. Interruption NDRT, 3. Hands SW and 4. Gaze Road). In a first step, to validate the model in general, the timestamps were plotted for each individual ($N = 14$) without distinguishing between the different NDRTs (Figure 4). As the figure shows, the variance of the measurement values is high and outliers can be detected. For statistical evaluation, median values rather than the arithmetic mean were used, as the median is more robust towards outliers (Schillinger, 2002). Data was tested on non-linearity, normal distribution, homoscedasticity and influential outliers using residual vs. fitted-, normal-Q-Q-, scale-location and residuals vs. leverage plots.

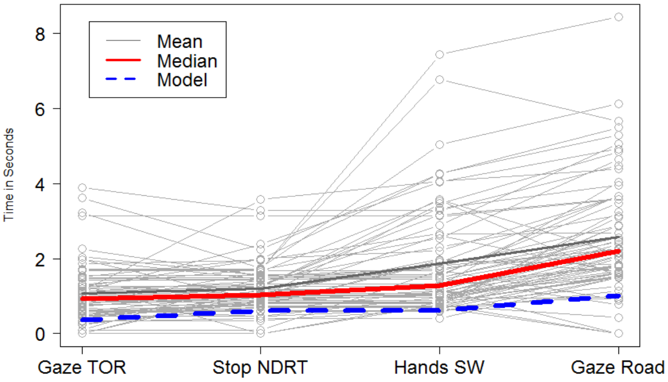

Figure 4: Course of the Participant Data and Model Prediction of Takeover Behavior Patterns ($N = 14$; Source: own figure).

Overall, the gaze to the TOR was performed with a median of 0.96s ($M = 1.07$s), ranging from a minimum (min) of zero seconds (s; zero seconds occur due to no NDRT condition, in which the gaze did not rest on a NDRT) to a maximum (max) of 3.3s. The NDRT was at the median of 1.08s interrupted ($M = 1.25$s, $min =$ 0s, $max = 3.58$s), the hands were after a median of 1.32s ($M = 1.91$s, $min = 0.66$s, $max = 7.44$s) at the steering wheel and after 2.28s ($M = 2.7$s, $min = 0$s, $max =$ 8.44s) the gaze was on the road. The model performs the sequence with the following times: (1) Gaze to TOR

Table 1: Correlations of NDRT with the model predictions and the correlation of accumulated empirical data with model predictions (significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1; source: own table).

NDRT	Correlation Coefficient and p-Value Arithmetic Mean	Correlation Coefficient and p-Value Median
Baseline	$r(2) = .96$, $p = .038$ *	$r(2) = .99$, $p = .006$ **
Listening	$r(2) = .96$, $p = .035$ *	$r(2) = .99$, $p = .012$ *
Searching	$r(2) = .86$, $p = .136$	$r(2) = .9$, $p = .1$
Reading	$r(2) = .8$, $p = .2$	$r(2) = .83$, $p = .17$
Tetris	$r(2) = .89$, $p = .1$	$r(2) = .95$, $p = .048$ *
Accumulated	$r(2) = .91$, $p = .086$ *	$r(2) = .95$, $p = .048$ *

0.37s, (2) Interruption of NDRT 0.62s, (3) Hands at SW 0.62s and (4) Gaze on Road 1.02s. It is noticeable, that the model is overall faster in the reaction times. As this applies for the overall time course, this difference is supposed to be justified by some patterns, that apply before the gaze is directed to the TOR. This is an important fact, that is further addressed in the discussion. It does not affect the overall evaluation though, as a time shift beforehand would shift the overall sequence. Statistical analysis of the data shows a significant bivariate pearsons correlation between the median behavior of the participants ($n = 14$) and the model predictions ($r(2) = .95$, $p = .05$). Figure 4 shows the course of the individuals over all takeover conditions, including the median course (bolt-red line), the mean course (dark-gray line) and the course of the model predictions (striped-blue line). The correlation shows that predictions of the model represent empirical data validly which strengthens proceeding model predictions of the action decision. In order to understand where the variances come from, further analysis concerning the different NDRTs is made. Table 1 shows that only NDRT conditions without additional physical movements show significant correlations with the cognitive model using mean as well as median values. These movements apply mainly for reading the newspaper (occupied hands) and searching something in the back (torso turned, hands occupied). The tetris on the mounted tablet also involves the hands, but no holding is necessary, hence the correlation is significant for median, but not for mean values. Overall, no significant correlation of model predictions can be found for conditions, in which additional motoric processes were necessary before taking over. This indicates, that the model can already display the cognitive occupation of NDRTs, but lacks to include motoric complex tasks. In further studies, additional cognitive processes will be investigated more detailed with eyetracking analysis.

Discussion

Results indicate, that the cognitive model is able to validly represent the time course during a takeover (a.). Only conditions, in which the hands are occupied by holding something (reading a newspaper) or turning the torso (searching something in the back), model predictions do not correlate significantly with empirical time courses. This is not a surprise however, as the ACT-R motor module does not yet include complex movements. This finding illustrates that as soon as the hands are occupied during a NDRT or the body is turned away, additional movement patterns are necessary (e.g. removing reading glasses, folding newspaper, turning body to front) before the defined takeover steps can be performed. Mechanisms that arise due to NDRTs can only partly be depicted (b.). These resource depending circumstances should be considered in the model for better time estimation of NDRT, depending on the occupied resource and NDRT cancellation criteria. As mentioned before, although significant correlations are found for the course of the takeover, the model is in its overall performance still faster. The cause of that may lie in the fact, that data was collected in real traffic environments. Thus, more environmental cues than the model includes are probably attended by participants before moving the gaze to the TOR (c.).

However, the current model provides a good basis for predictions of the takeover. It still is a general model and does not depict individual differences. The model will be refined next, incorporating the impact of the complexity of a situation. Eye movement patterns will be captured for more particularized understanding of processes in the vision module. Along with this, further development of the SA update as well as of the action decision will be validated. The current cognitive model is going to be expanded to enable the prediction of individualized cognitive processes depending on the situation complexity. These predictions will serve as groundwork for further specification of cognitive processes in complex traffic situations and resulting action decisions, that should be linked to eye-movement patterns. Resulting predictions of human cognition consequently serve for the adaption of HMI concepts.

Conclusion

The present model allows the prediction of takeover patterns in highly automated driving. It is able to illustrate several steps that are undertaken during the takeover process validly. This approach provides a solid groundwork for a more specific development of the model. The impact of complexity aspects on cognitive processes during the takeover will further be included. In the next step, research on complex situations will be done. To understand how the cognitive system and especially the visual perception handles complex situations, eye tracking data will be gathered. By implementing and comparing this approach to empirical data the underlying cognitive processes can be elucidated and resulting necessities during a takeover understood. This will later serve as guideline for the development of suitable HMI systems.

Acknowledgements

I wish to acknowledge the help provided by my supervisors Michael Schulz and Kathrin Zeeb from Robert Bosch GmbH and the public promoted projects PAKoS and Ko-HAF.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036.

Bellet, T., Bornard, J.-C., Mayenobe, P., Paris, J.-C., Gruyer, D., & Claverie, B. (2012). Computational simulation of visual distraction effects on car drivers'situation awareness. In *Iccm 2012*.

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2015). What makes interruptions disruptive?: A process-model account of the effects of the problem state bottleneck on task interruption and resumption. In *Proceedings of the 33rd annual acm conference on human factors in computing systems* (pp. 2971–2980). ACM.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human factors*, *37*(1), 32–64.

Feldhütter, A., Gold, C., Schneider, S., & Bengler, K. (2017). How the duration of automated driving influences take-over performance and gaze behavior. In *Advances in ergonomic design of systems, products and processes* (pp. 309–318). Springer.

Gold, C., Damböck, D., Lorenz, L., & Bengler, K. (2013). "take over!" how long does it take to get the driver back into the loop? In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 57, *1*, pp. 1938–1942). SAGE Publications Sage CA: Los Angeles, CA.

Ko-HAF. (2017). Ko-haf - wizard-of-oz-konzept. YouTube. Retrieved from https://www.youtube.com/watch?v=4mm3xaBfQZc

Merat, N., Jamson, A. H., Lai, F. C., Daly, M., & Carsten, O. M. (2014). Transition to manual: Driver behaviour when resuming control from a highly automated vehicle. *Transportation research part F: traffic psychology and behaviour*, *27*, 274–282.

Naujoks, F., & Neukum, A. (2014). Timing of in-vehicle advisory warnings based on cooperative perception. In *Proceedings of the human factors and ergonomics society europe chapter annual meeting* (pp. 193–206). HFES Torino.

Ratwani, R. M., McCurry, J. M., & Trafton, J. G. (2010). Single operator, multiple robots: An eye movement based theoretic model of operator situation awareness. In *Proceedings of the 5th acm/ieee international conference on human-robot interaction* (pp. 235–242). IEEE Press.

SAE, T. (2014). *Surface vehicle information report. taxonomy and definitions for terms related to on-road motor vehicle automated driving systems.* SAE International.

Salvucci, D. (2006). Modeling driver behavior in a cognitive architecture. *Human factors*, *48*(2), 362–380.

Salvucci, D., Boer, E., & Liu, A. (2001). Toward an integrated model of driver behavior in cognitive architecture. *Transportation Research Record: Journal of the Transportation Research Board*, (1779), 9–16.

Schillinger, M. P. (2002). *Flächenhafte schätzung mit classification and regression trees und robuste gütebestimmung ökologischer parameter in einem kleinen einzugsgebiet* (Doctoral dissertation).

Taatgen, N. (1999). The atomic components of thought. *Trends in Cognitive Sciences*, *3*(2), 82.

Walch, M., Lange, K., Baumann, M., & Weber, M. (2015). Autonomous driving: Investigating the feasibility of car-driver handover assistance. In *Proceedings of the 7th international conference on automotive user interfaces and interactive vehicular applications* (pp. 11–18). ACM.

Zeeb, K., Buchner, A., & Schrauf, M. (2015). What determines the take-over time? an integrated model approach of driver take-over after automated driving. *Accident Analysis & Prevention*, *78*, 212–221.