The influence of vestibular feedback on transitions between different levels of automation

Pia Wald[1], Laura Hiendl[2], Martin Albert[3], & Klaus Bengler[1]
[1]Technical University of Munich, [2]University of Regensburg, [3]AUDI AG, Germany

Abstract

The driver's tasks and responsibilities vary in a multi-level automated driving car. While drivers have to monitor the system and the environment in assisted and partially automated driving, they can engage in non-driving related tasks during higher levels of automation. To support drivers in their tasks and increase their mode awareness, the system should provide comprehensible feedback about it's state and intentions. Two different feedback concepts were implemented for this purpose, comparing a visual-auditory with a visual-auditory-vestibular feedback. A driving study (*N*=47) was conducted with a test vehicle simulating partially and highly automated motorway driving. Depending on their experience with adaptive cruise control (ACC), participants were split into three groups and experienced manual, partially and highly automated driving as well as transitions between these levels. The results revealed that both concepts generated high levels of trust and acceptance. Experience with ACC showed no significant effect. However, visual-auditory feedback with additional vehicle motions could significantly increase the predictability of the automated vehicle's behaviour. Moreover, in partially automated driving visual-auditory-vestibular feedback was perceived as more relieving than without vehicle motions.

Introduction

The driver's role is changing as automated driving functions become increasingly widespread. According to the taxonomy of the Society of Automotive Engineers (SAE, 2016), automated driving vehicles (SAE L2-L5) can perform both lateral and longitudinal vehicle guidance, with only the driver's responsibilities changing. During partially automated driving (SAE L2), the driver has to monitor the automated system and the environment. In higher levels of automation (LoA), the driver is allowed to withdraw from supervising and accomplish a non-driving related task (NDRT). Future vehicles may combine several LoA. The greatest challenges facing these multi-level systems are not only the variation in responsibility for the driving task, but also the transitions between different LoA. Literature to date has mainly considered questions regarding the time until manual control is regained and the respective influencing factors (Zhang et al., 2019). The available time budget and driver's reaction to a take-over request (TOR) mostly range from five to ten seconds (e.g., Gold et al., 2013), whereas studies regarding the mental stabilization time after a transition have shown

In D. de Waard, S.H. Fairclough, K.A. Brookhuis, D. Manzey, L. Onnasch, A. Naumann, R. Wiczorek, F. Di Nocera, S. Röttger, and A. Toffetti (Eds.) (2022). Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2022 Annual Conference. ISSN 2333-4959 (online). Available from http://hfes-europe.org

that it takes the driver up to 40 seconds to regain full attention (Merat et al., 2014). This means that the cognitive processing of a take-over situation takes more time than the (reflexive) motoric response to a TOR (Zeeb, 2016) and should be further addressed (Merat et al., 2014).

The varying responsibilities (SAE, 2016) in a multi-level system and transitions leads to new challenges in the HMI design (Othersen, 2016). Regular feedback from the vehicle is important to keep the driver informed, despite the more passive role. Feedback can be provided in a variety of ways and via all modalities (Bengler et al., 2020). As state of the art, feedback is usually presented visually (Albert et al., 2015), auditorily (Forster et al., 2017) or tactilely (Petermeijer et al., 2017). The design and information content of feedback depends on the respective LoA and thus on the driver's task (Beggiato et al., 2015; Bengler et al., 2020). In this context, multimodal feedback and interfaces are advantageous, resulting in a better system awareness (Bengler et al., 2020; Wickens, 2002). Moreover, multimodal feedback can be complemented by active vehicle motions covering the vestibular modality. These motions provide a new possibility to communicate intentions of the automation before initiating a manoeuvre. Previous studies showed for partially automated driving that detecting a preceding vehicle should be announced via pitch motions (Cramer et al., 2018). Additionally, roll motions should announce lane changes (Cramer, 2019). These pitch and roll motions have been considered useful in assisting drivers regarding their mode and system awareness (Cramer et al., 2018; Cramer, 2019).

There is still some uncertainty about vestibular feedback in a multi-level automated vehicle. The aim of this study is to examine whether additional vestibular feedback can improve the driver's mode awareness. Furthermore, based on previous results showing a correlation between experience with adaptive cruise control (ACC) in years and evaluation of feedback (Wald et al., 2021), the influence of experience with ACC on the assessment of the feedback concepts will be investigated. Additionally, this study provides an overview of activation times for different transitions depending on feedback. As most of the transition studies so far took place in driving simulators, those findings have to be confirmed in a real road environment (Zhang et al., 2019) and need to include more realistic scenarios (Eriksson & Stanton, 2017). Thus, two different feedback concepts, one with active vehicle motions and one without motions, were investigated in a real-world driving study with uncritical transitions between different LoA by three groups depending on the ACC experience.

Method

Sample

Forty-seven drivers with a mean age of 32.91 (*SD* = 9.93) years, ranging from 22 to 59, participated in the experiment. The sample represented a variation of gender and technical background (23.4% technical female, 25.6% non-technical female, 31.9% technical male and 19.1% non-technical male). Participants drove an average of 14,468 km a year (*SD* = 8,888 km) before and 9,000 km a year (*SD* = 5,782 km) during the period influence by the COVID-19 pandemic situation. 72% of the sample had used lane keeping assistance and 36% partially automated driving systems before. Moreover, 66% had previous experience with ACC, 16 participants with little experience (*M* = 1.29, *SD* = 0.77, *min* = 0.1, *max* = 2) and 15 with high experience (*M* = 8.73, *SD* = 4.53, *min* = 3, *max* = 18) with ACC.

Test setup and equipment

The driving study was conducted on the three-lane German A9 motorway between the Manching and Denkendorf exits. However, only the right and middle lane of the motorway was used for safety reasons. The test vehicle, an Audi A5 (year of construction 2012), drove at a maximum of 120 km/h. A prototypical automation system was implemented that was able to simulate partially (SAE L2) and highly (SAE L4) automated driving. The test vehicle performed lateral and longitudinal vehicle guidance.

The participant sat in the driver's seat, and there were two further experimenters in the test vehicle. The experimenter in the passenger seat acted as a safety driver. Additional equipment such as a second interior mirror, additional exterior mirrors, driving school pedals, and a monitor to display essential information about the system, assisted the safety driver. Besides triggering lane changes, this experimenter could adapt to speed limits and provoke HMI elements, pitch and roll motions (referring to Cramer et al., 2018; Wald et al., 2021). The second experimenter sat in the back seat, coordinated the questionnaires and gave the participants instructions.

Study procedure

The driving study took place during the COVID-19 pandemic situation, so a hygiene concept was developed with experts beforehand which is similar to Wald et al. (2021). Fig. 1 presents the sequence of the driving study. The experiment was conducted in German. Participants initially received verbal instruction on the procedure, the test vehicle operation, and the various transitions (Figure 1). They then practised activating the different LoA in the stationary test vehicle, which was followed by the test drives. During all driving sessions, participants drove manually on the motorway and activated the automation system in the right lane. During the first three minutes of the settling-in drive, the test vehicle performed no lane changes since the participants got familiar with the system. Drivers received neither visual nor vestibular feedback during the settling-in phase, but only basic information such as current speed and position in the instrument cluster. Subsequently, they experienced two feedback concepts consisting of four transitions in a randomized order. During L4, participants had to play a game on a tablet mounted in the centre console.

Human-Machine Interface

The human-machine interface consisted of visual elements in the instrument cluster, auditory signals, and active vehicle motions. According to literature recommendations (Beggiato et al., 2015), system's status, future and current manoeuvres, current velocity and a preceding vehicle were presented in the cluster (Wald et al., 2021). The LoA were displayed in different colours (L2 in blue, L4 in green) for this driving study. Lane changes were announced with an arrow in the cluster, the direction indicator and additional active roll motions in the vestibular concept. A degressive roll profile with an angle of 3.0° and an acceleration of -4.5°/s² was used to announce lane changes (Cramer, 2019; Wald et al., 2021). Moreover, pitch motions announced a slower detected preceding vehicle with an angle of 1° and an acceleration of $-5°/s^2$ (Cramer, 2019).

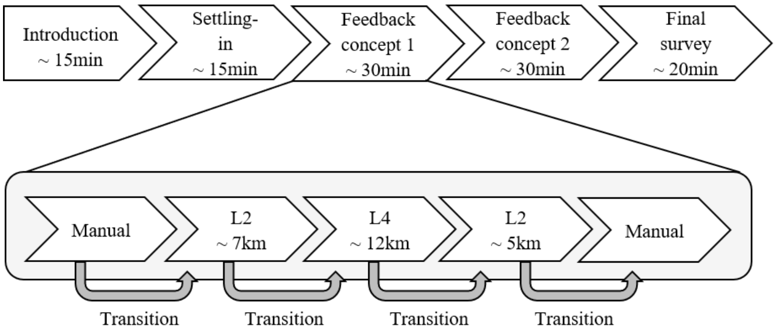

Figure 1. Sequence of the driving study and transitions.

Immediately after entering the motorway, a system suggestion to activate L2 was displayed. Based on experts' advice for an unobtrusive suggestion, no additional auditory hint was given. However, a sound announced the other transitions for the following LoA with an additional pop-up. The transition pop-up in L2 indicated that L4 was available, whereas a transition pop-up in L4 requested the driver to activate L2. A task description was shown after the transition had been accepted. Depending on the following LoA, the description indicated that performing a NDRT (in L4) was allowed or prompted the driver to fully monitor the system (in L2). A transition to manual driving (SAE L0) was announced with red symbols and an intrusive sound.

Processing and evaluation of the data

Objective data included vehicle data and internal data from the automation system, from which *activation times* were calculated. After each transition, questions were asked about mode awareness. At the end of both concept driving parts, the participants answered questionnaires regarding their subjective perception of the feedback concepts. *Trust* in automation was assessed by the questionnaire from Körber (2019) which is divided into six subscales on a five-point Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree"). Participants rated the three subscales Reliability/Competence, Understanding/Predictability, and Trust in Automation to obtain respective trust of each feedback concept. The German version of the questionnaire by van der Laan et al. (1997) was used to evaluate *acceptance* of the feedback concepts. This survey is divided into the subscales usefulness and satisfying based on nine items on a five-point scale from -2 to 2. *Mode awareness* was measured with two questions (Othersen, 2016) after each transition for the previous mode on a 15-point scale consisting of five categories from "very little" to "very strong" with the additional opportunity "no answer". Participants were asked to orally validate their *task awareness* ("I was always aware which tasks I had and which ones the system had.") and their *monitoring behaviour* ("I have permanently monitored the system."). After each concept drive, participants were asked to rate specific statements for the *feedback characteristics* on a 7-point rating scale from 1 ("does absolutely not apply") to 7 ("does absolutely apply"). Three statements, each for L2 and L4, stated whether the feedback was perceived as annoying, distracting and relieving. Moreover, predictability of the automated vehicle was validated after each transition with the statement "How predictable was the system behaviour in the previous mode?" (Petermann-Stock, 2015) on a 15-point scale.

The data were analysed using MATLAB, the statistics with R. For this study, a repeated measure mixed design was used combining the between-subject factor experience with ACC and depending on the dependent variable, the within-subject factors feedback concept, transition and LoA. The participants received both visual-auditory (VA) and visual-auditory-vestibular (VAV) feedback. The sample was divided by their experience with ACC (zero, little and high). An analysis was performed and interpreted, even if the Shapiro-Wilk test showed significance, as the ANOVA is considered robust against a violation of the normal distribution (Blanca et al., 2017). A significance level of $\alpha = 0.05$ was initially applied and partial eta-squared was computed as effect size statistics. Degrees of freedom were corrected when Mauchly's test for sphericity showed significance (Greenhouse-Geisser). Homogeneity of variance was assessed by Levene's test for equality of error variances and homogeneity of covariances was calculated by Box's test for equality of covariance matrices. Unless otherwise stated, data was homogenous in variance and covariance. Post hoc comparisons were controlled with Benjamini-Hochberg corrected p-values (Benjamini & Hochberg, 1995).

Results

Activation times

The values for the mean (*M*), standard deviation (*SD*) as well as minimum and maximum for each transition depending on the feedback concept are presented in Table 1. Analysis of variance for activation time finds neither a significant effect of experience with ACC or feedback, nor any significant interactions ($p > .05$). However, ANOVA yielded a significant main effect for transition ($F(1.62,63.18) = 18.48$, $p < .001$, $\eta_p^2 = 0.321$). Following post-hoc analysis revealed that the activation time for transition from L4 to L2 ($M = 8.77$, $SD = 4.83$) is significantly higher compared to transition from L2 to L4 ($M = 5.33$, $SD = 2.26$, $p < .001$) and to transition from L2 to L0 ($M = 5.64$, $SD = 3.3$, $p < .001$).

Table 1. Descriptives of participants' activation times for different transitions depending on feedback concept.

	Visual-auditory			Visual-auditory-vestibular		
	M (SD)	*Min*	*Max*	*M (SD)*	*Min*	*Max*
L2 to L4	5.41 (2.66)	2.36	15.77	5.24 (1.81)	2.82	12.68
L4 to L2	8.48 (3.7)	3.54	24.99	9.07 (5.77)	2.86	31.48
L2 to L0	5.34 (3.47)	1.71	23.78	5.95 (3.13)	2.26	15.96

Trust and acceptance

Trust is presented in Figure 2. Overall, both concepts were evaluated as reliable, predictable, and generated high trust in automation. The applied ANOVA indicated no significant differences between the feedback concepts for Reliability ($F(1,44) = 1.79$, $p = .188$, $\eta_p^2 = 0.039$), Predictability ($F(1,44) < 1$, $p > .05$) and Trust in Automation ($F(1,44) = 1.27$, $p = .267$, $\eta_p^2 = 0.028$). Moreover, there was neither an effect of experience with ACC nor an interaction between experience and feedback for all three subscales ($p > .05$).

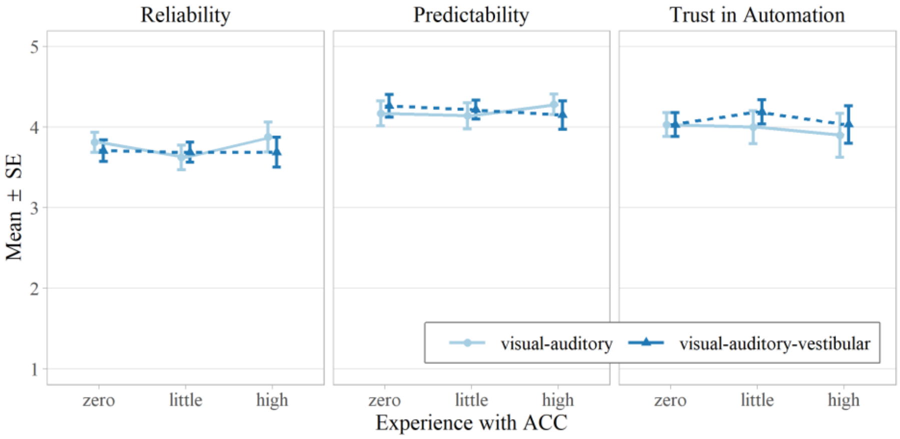

Figure 2. Participants' mean ratings of the feedback concepts for the three dimensions of the questionnaire from Körber (2018) depending on the experience with ACC.

Concerning acceptance, both concepts were rated as useful (VA: $M = 0.74$, $SD = 0.32$, VAV: $M = 0.74$, $SD = 0.37$) and satisfying (VA: $M = 1.38$, $SD = 0.48$, VAV: $M = 1.34$, $SD = 0.64$). However, analyses of variance showed neither significant differences between the feedback concepts ($F(1,44) < 1$, $p > .05$) nor between the experience with ACC ($F(2,44) < 1$, $p > .05$) for both scales. Moreover, there were no interaction effects between feedback and experience for either scale.

Mode Awareness

Results for task awareness showed that neither experience with ACC ($F(2,44) = 1.89$, $p = .163$, $\eta_p^2 = 0.079$) nor the feedback concept ($F(1,44) < 1$, $p > .05$) yielded a significant effect. However, there was a significant effect of LoA ($F(1.77,77.68) = 17.34$, $p < .001$, $\eta_p^2 = 0.283$). Post hoc tests revealed a significant higher task awareness for L0 ($M = 14.37$, $SD = 1.28$) compared to the first ($M = 12.67$, $SD = 2.67$, $p < .001$) and the second ($M = 12.9$, $SD = 2.09$, $p < .001$) L2 section. Moreover, L4 ($M = 13.85$, $SD = 1.85$) generated a higher task awareness than the two L2 sections ($p < .001$). The applied ANOVA indicated no significant interaction effects. Figure 3 presents the result.

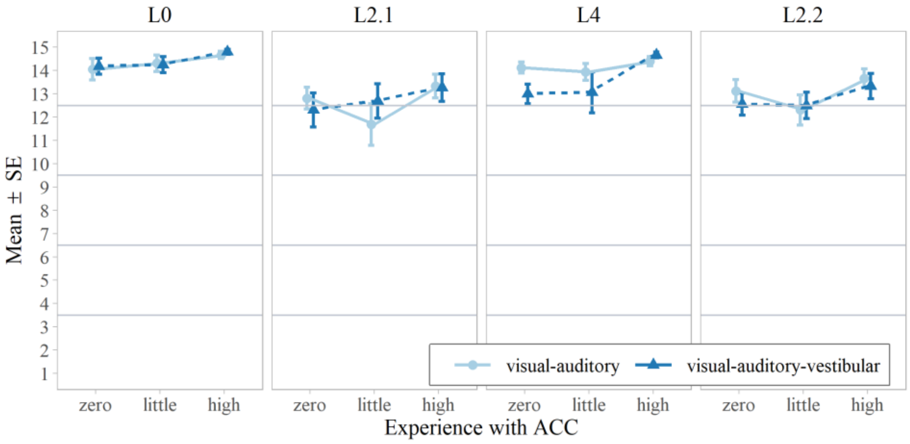

Figure 3. Participants' mean ratings for the feedback concepts of their task awareness depending on the experience with ACC.

Additionally, participants should monitor the system constantly during L2 and not at all during L4. The ANOVA for self-rated monitoring behaviour for the two L2 sections revealed neither a main effect for experience with ACC ($F(2,44) < 1$, $p > .05$) nor for feedback ($F(1,44) < 1$, $p > .05$). However, a significant effect of LoA was found ($F(1,44) = 6.73$, $p = .013$, $\eta_p^2 = 0.133$), indicating a decreasing monitoring behaviour from the first ($M = 11.4$, $SD = 2.57$) to the second ($M = 10.69$, $SD = 2.87$) L2 section. Furthermore, a significant interaction between LoA and feedback ($F(1,44) = 4.54$, $p = .039$, $\eta_p^2 = 0.093$) was noted. Subsequent post hoc analysis showed no significant differences ($p > .05$). Moreover, there were no further significant interaction effects. The applied ANOVA for L4 revealed no significant effects for experience with ACC ($F(2,44) = 1.27$, $p = .292$, $\eta_p^2 = 0.054$) and feedback ($F(1,44) < 1$, $p > .05$). The interaction between the two factors achieved statistical significance ($F(2,44) = 3.28$, $p = .047$, $\eta_p^2 = 0.013$), but post hoc tests showed no significant results.

Feedback characteristics

The mean values for distracting, annoying and relieving can be found in Table 2. Analysis of variance for annoying and distracting found neither a significant main effect nor any significant interactions ($p > .05$). On a descriptive level, VA seems to be less annoying and less distracting in L4 (cf. Table 2). The ANOVA for relieving yielded no significant differences in experience, LoA or feedback. However, the interaction between feedback and LoA reached statistical significance ($F(1,44) = 5.37$, $p = .025$, $\eta_p^2 = 0.109$). Post-hoc comparisons indicated that VAV is more relieving than VA ($p = .006$) in L2.

Table 2. Assessment of feedback characteristics for L2 and L4 depending on the feedback concept.

		Distracting		*Annoying*		*Relieving*	
		M	*SD*	*M*	*SD*	*M*	*SD*
L2							
	Visual-auditory	1.70	0.81	1.40	0.68	4.43	1.64
	Visual-auditory-vestibular	1.70	0.91	1.40	0.68	5.04	1.35
L4							
	Visual-auditory	1.81	1.48	1.49	1.04	4.53	2.14
	Visual-auditory-vestibular	2.26	1.67	1.81	1.42	4.40	2.13

Concerning predictability, analysis of variance found neither a significant effect for experience with ACC nor for LoA or any significant interaction. Feedback, however, had a significant effect on the predictability ($F(1,41) = 5.77$, $p = .021$, $\eta_p^2 = 0.123$). Participants rated VAV ($M = 11.57$, $SD = 2.49$) as more predictable than VA ($M = 11.21$, $SD = 2.73$).

Conclusion and Discussion

The aim of the current study was to examine whether additional vestibular feedback can improve driver’s mode awareness. Therefore, two different feedback concepts for a multi-level system with partially and highly automated motorway driving were evaluated. In general, both concepts generated high trust and acceptance scores. These results are consistent with previous research (Cramer, 2019; Wald et al., 2021). Moreover, results for the single item predictability revealed that VAV was more predictable compared to VA, although the subscale Understanding/Predictability of the trust questionnaire showed no differences between the feedback concepts. This inconsistency may be due to the fact that the subscale considered understanding in addition to predictability, thus allowing a more precise measurement. Furthermore, VAV was more relieving in L2. These results support the findings of Cramer (2019). However, additional active vehicle motions in L4 appear to be distracting and annoying on a descriptive level. Contrary to expectations, this study did not find a significant difference between the concepts for task awareness and monitoring behaviour. Results revealed that L2 generated a lower task awareness than L0 and L4. These findings further support the idea of recent studies indicating that the driving task should either be fully undertaken by the driver or completely surrendered to the automated driving system (Petermann-Stock, 2015). Additionally, this study found that the monitoring behaviour decreased after L4. Activation times for the transition from L4 to L2 were higher compared to other transitions (L0 to L2 and L2 to L0) in uncritical situations. This result may be explained by the fact that drivers had to deflect from the NDRT and orientate themselves in the environment. Surprisingly, no differences were found between the experience in ACC which is contrary to a previous study (Wald et al., 2021).

The generalisability of these results is subject to certain limitations. Due to the real-world scenario, standardisation of the requirements is difficult since the surrounding traffic and the weather are not controllable. To ensure similar conditions, the study proceeded on same times during the day. Moreover, participation in the study was voluntary, what might have positively influenced the results because participants were interested in automated driving. Overall, this study strengthens the idea that additional vestibular feedback can improve the human driver interaction in partially automated driving. Based on these and previous results, further research should combine different feedback strategies in a multi-level system (e.g., using vestibular feedback only in partially automated driving) to support drivers in their tasks.

Acknowledgments

We would like to thank our colleagues, in particular Stephan Bültjes and Stephanie Cramer for their assistance with the test vehicle hardware as well as the software. The Ethics Board of the Technical University Munich provided ethical approval for the hygiene concept and this study, the corresponding ethical approval code is 295/21 S.

References

Albert, M., Lange, A., Schmidt, A., Wimmer, M., & Bengler, K. (2015). Automated Driving - Assessment of Interaction Concepts Under Real Driving Conditions. In *6th International Conference on Applied Human Factors and Ergonomics and the Affiliated Conferences, AHFE 2015.*

Beggiato, M., Hartwich, F., Schleinitz, K., Krems, J., Othersen, I., & Petermann-Stock, I. (2015). What would drivers like to know during automated driving? Information needs at different levels of automation. In *7. Tagung Fahrerassistenz.*

Bengler, K., Rettenmaier, M., Fritz, N., & Feierle, A. (2020). From HMI to HMIs: Towards an HMI Framework for Automated Driving. *Information*, *11*.

Benjamini, Y., & Hochberg. Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, 57,* 289–300.

Blanca, M.J., Alarcón, R., Arnau, J., Bono, R., & Bendayan, R. (2017). Non-normal data: Is ANOVA still a valid option? *Psicothema*, *29,* 552-557.

Cramer, S. (2019). *Design of Active Vehicle Pitch and Roll Motions as Feedback for the Driver During Automated Driving*. PhD thesis. Technische Universität München.

Cramer, S., Kaup, I., & Siedersberger, K.-H. (2018). Comprehensibility and Perceptibility of Vehicle Pitch Motions as Feedback for the Driver During Partially Automated Driving. *IEEE Transactions on Intelligent Vehicles*, *4*, 3-13.

Eriksson, A., & Stanton, N.A. (2017). Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control. *Human Factors, 59,* 689-705.

Feldhütter, A., Segler, C., & Bengler, K. (2018). Does Shifting Between Conditionally and Partially Automated Driving Lead to a Loss of Mode Awareness? In Stanton N. (Eds), *Advances in Human Aspects of Transportation. AHFE 2017*. Springer

Forster, Y., Naujoks, F., & Neukum, A. (2017). Increasing anthropomorphism and trust in automated driving functions by adding speech output. In *IEEE Intelligent Vehicles Symposium 2017* (pp. 365-372).

Gold, C., Damböck, D., Lorenz, L.M., & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 57,* 1938-1942.

Körber, M. (2019). Theoretical Considerations and Development of a Questionnaire to Measure Trust in Automation. In Bagnara S., Tartaglia R., Albolino S., Alexander T., Fujita Y. (Eds*) Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018).*

Merat, M., Jamson, A.H., Lai, F.C.H., Daly, M., & Carsten, O.M.J. (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.

Othersen, I. (2016). *Vom Fahrer zum Denker und Teilzeitlenker*. PhD thesis. Technische Universität Braunschweig.

Petermann-Stock, I. (2015). *Automation und Transition im Kraftfahrzeug*. PhD thesis. Technische Universität Braunschweig.

Petermeijer, S.M., Cieler, S., & de Winter, J.C.F. (2017). Comparing spatially static and dynamic vibrotactile take-over requests in the driver seat. *Accident Analysis & Prevention, 99,* 218-227.

SAE. (2016). *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* (2016-09 ed.) (No. J3016).

Van der Laan, J.D., Heino, A., & De Waard, D. (1997). A Simple Procedure for the Assessment of Acceptance of Advanced Transport Telematics. *Transportation Research Part C: Emerging Technologies*, *5*, 1–10.

Wald, P., Haentjes, J., Albert, M., Cramer, S., & Bengler, K. (2021). Active Vehicle Motion as Feedback during Different Levels of Automation. In *IEEE International Intelligent Transportation 2021* (pp. 1713–1720).

Wickens, C.D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3,* 159-177.

Zeeb, K., Buchner, A., & Schrauf, M. (2016). Is take-over time all that matters? The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. *Accident Analysis and Prevention, 92,* 230-239.

Zhang, B., de Winter, J.C., Varotto, S., Happee, R., & Martens, M.H. (2019). Determinants of take-over time from automated driving: A meta-analysis of 129 studies. *Transportation Research Part F: Traffic Psychology and Behaviour, 64,* 285-307.