Automation in Railway Operations: Effects on Signaller and Train Driver Workload

Niels Brandenburger, Birte Thomas-Friedrich, Anja Naumann, Jan Grippenkoven
German Aerospace Center (DLR e.V.)
Institute of Transportation Systems
Lilienthalplatz 7
38108 Braunschweig
niels.brandenburger@dlr.de

Abstract:

Throughout the railway domain, increasing levels of automation are employed to ensure safety and increase efficiency on the tracks. This impacts the task characteristics of the signaller and train driver. In a scenario of German railway automation, an automated interlocking system routes trains automatically and trains equipped with "ATO over ETCS" ("Automatic Train Operation over the European Train Control System") automatically drive along the predefined routes adhering to speed restrictions. Thus, the task load of signallers and train drivers decreases, as manual inputs decline in favour of monitoring the functioning of automatic systems. Yet, it remains unclear whether decreasing task load directly lowers subjective workload. Additionally, the question of optimal workload levels has yet to be addressed. In order to ensure optimal performance, under- and overload are to be avoided. In previous studies on workload in railway operations, workload was assessed without considering an individual optimum of workload. In the simulator studies described in the paper, subjective workload was assessed with the goal of analysing the impact of automation on achieving an optimal workload level. In two separate studies with train drivers and signallers, subjective workload was assessed by two self-report measures (NASA-TLX and DLR-WAT) after participants completed periods of manual driving and driving with automated systems. Results consistently indicate lower subjective workload in the automated work settings for both signallers and train drivers. Interestingly, signaller workload was close to optimal subjective levels while train driver workload scores were considerably lower than optimal. This highlights the need for thoughtful introduction of automation into the train driver environment. Furthermore, the DLR-WAT differentiates between mental workload caused by the different stages of information processing. In train drivers and signallers, workload stemming from information perception seems to be more pronounced than workload stemming from mental operations occurring at later stages of information processing, especially in automatic work settings. Assessing workload relative to an individual optimum and differentiating causes of mental workload along the different stages of information processing offers unique insights into signaller and train driver workload. The results make it possible to ascertain, which specific aspects of the introduction of automation in the signaller and train driver tasks lead to lowered overall workload.

Introduction

Numerous studies have pointed out efforts in the railway domain to increase capacity and efficiency through increased automation [1] making use of digital data available from a variety of sources [2][3][4]. Introductions of advanced technology inevitably change the roles that humans play as part of the railway system. While new approaches to integrate systems with high levels of automation [5][6] into existing railway operational structure are being discussed [7][8], the human workplaces in the railway domain have already undergone a remarkable transition with the past decades. Railway operation today already relies on very different task allocations between technology and human than thirty years ago. For formalizing the task allocation between technology or automation and humans, the well cited model incorporating levels and stages of automation by Parasuraman, Sheridan and Wickens [5] proved as a useful tool to track the development. The assumption behind the model is that human information processing can be simplified as a sequential four stage process. Information is basically sensed and perceived; then a decision is made based on that information and that decision is implemented into an action [5]. Based on that assumption, the model defines any automated system as having a certain level of automation ranging from *no automation present* to *autonomous automation* on each of the four stages of human information processing [5]. We can use this model to describe the incremental changes in different occupational settings in the railway domain.

In case of the train driver, in- cabin signalling as an early step of increasing automation of information acquisition, has fundamentally changed the primary focus of attention of the train driver along with the required behaviour [9][10][11] and automatic train protection systems (ATP) as for example national train control systems (NTC) like the German "Punktförmige Zugbeeinflussung" (PZB) automatically prevent certain violations. This early step of automation is especially pronounced since automatic train protection systems featuring in-cabin signalling are equipped with additional features interfering with the train driver´s choice of speed, if deemed unsafe. Technological solutions for providing accurate data about positions, maximum speeds and movement authorities in the cabin though digitalization, as does for example the "European Train Control System" (ETCS), further altered the spectrum of tasks of the train driver. It can be argued that with in-cabin signalling, the level of automation increased and changed train driving with regard to the early stages of information processing. As a negative consequence of this change, train drivers reduce the time spent in attending world outside the cab [9][10]. In terms of the model by Parasuraman, Sheridan and Wickens [5], we can further describe ATP systems as an increase in levels of automation at the latter stages of information processing, namely the stages of decision making and action implementation, because technology decides whether a certain speed is seen as safe and acts if deemed unsafe. Naweed concludes that "these systems and approaches have introduced increasing layers of autonomous control in the train" [9]. Similarly, in the signaller workplace, the automatic route setting function called "Zuglenkung" (ZL) has already altered the signaller`s daily work as signal aspects. For example points are manipulated depending on the scheduled train without manual input from the signaller [12]. This development is an increase in the level of automation in the stages of decision making and action implementation, the last stage of information processing.

Analysing the consequences of the substantial changes in human tasks in the railway system that result from the introduction of increasingly automated systems, we need to focus on the interplay between operators and technology to ensure safe and efficient operations. On the train driver side, automatic train operation (ATO) is the most important technological and operational challenge about to happen that needs to be rolled out into mainline service on the basis of sound considerations. Automatic train operation is defined roughly based on the railway specific grades of automation [13]. A more detailed translation into the concept of levels and stages of automation by Parasuraman, Sheridan and Wickens [5] is provided by [14]. The main characteristic of ATO is the automatic detection of active speed restrictions and the automatic adjustment of train speed accordingly taking into consideration braking curves, movement authorities and scheduled stops. Hence, ATO leads to further increased levels of automation at the later stages of information processing, since decisions are taken over by the systems and necessary actions are implemented by the automated system continuously. In a probable scenario of future automated railway operation, routine operations are handled automatically on both sides of railway operation, the track side and the rolling stock side. Automatic route setting assigns routes to the rolling stock ensuring safe positioning of the switches and appropriate signalling, while the rolling stock is controlled by ATO systems on every vehicle making sure a particular vehicle is driving at the correct speed at any given position. Considering the implications of this scenario for the signaller´s and the train driver´s tasks, there are major similarities for both occupations. The most prominent similarity is that for both positions manual task load is decreasing drastically in routine operations [15][16][12]. Signallers do not set routes manually if the automatic route setting is working without conflict and likewise train drivers do not adjust speeds of vehicles while ATO is active. Due to the decrease in task load, especially in the stage of manual task execution, the supervision of the system state will yet again evolve as the central task of operators in the future railway system. This means that continuous visual monitoring of automatically executed actions becomes most relevant. In light of the changes in task allocation the question arises how operator workload will be affected in this future scenario and whether drivers and signallers will still be able to execute their visual monitoring tasks reliably. Thus, we are ultimately interested in the link between the new vigilance tasks at hand and the performance achieved. Obviously, task requirements do not directly translate into a certain level of performance. Depending on the individual characteristics of an operator, a certain level of task load may translate into very different subjective levels of workload, which in turn plays a critical role in how well certain tasks are being performed [17][18][19]. Young et al. point out the difficulty of differentiating "the interaction between cognitive workload and physical workload" and their effects on performance [17]. As we highlight a shift from manual to mental tasks in the railway domain, we need to take into account possible interactions between sources of workload. Workload and the subdomain of mental workload can be conceptualized as a continuum ranging from underload to overload at the far ends of the continuum [19]. It is assumed that intermediate medium level of mental workload is most beneficial for performance outcomes [19][20]. Extreme underload is associated with deactivation, task-induced mental fatigue and drowsiness [18], while overload goes along with pronounced spenditure of effort and stress leading to degradation of performance [20]. Concerning the train driver, research has already shown poorer performance and slower reaction times to critical events in this scenario of high automation featuring ATO [14][21]. This finding replicates findings from other domains stating that increased levels of automation and lower workload at the latter stages of information processing, especially action implementation, result in poorer operator performance during irregular operational situations [22]. Some conclusions go even further suggesting inadequate workload to be causally linked to subsequent accidents [23]. Yet, it remains unclear how decreasing manual task load in combination with the remaining task load stemming from visual monitoring in our scenario actually affects workload in the context of signalling and train driving. This lack of clarity about workload development comes from two areas.

First, it concerns the direction of the deviation of workload from an optimal intermediate level. Some research suggests that increased automation of decision making and action implementation leads to low overall workload [22][6]. This implies a deviation from an optimal workload level towards underload. Other research claims that additional effort needed to sustain prolonged attention for visual monitoring actually leads to higher overall workload [11][24], thus suggesting overload through task- related effort [19]. To clarify these conflicting results for our scenario of future railway operation, we aim to establish the direction and magnitude of the deviation of overall workload from an optimal intermediate level in signallers and train drivers through the DLR- WAT [25] workload questionnaire. The DLR- WAT requests an indication of the workload deviation from an explicitly depicted optimal level. Deviations from an optimal level of workload could point to fields of future interventions in order to prevent operator performance decrements and to ensure safe operations. Secondly, the workload of signallers and train drivers will be affected differently at different stages of information processing in our scenario of future railway operation. We need to define at which stages of information processing signaller and train driver workload is especially vulnerable for deviations from an optimal level. Therefore, we need to investigate sub dimensions of mental workload relating to the stages of information processing and pinpointing specific stages of cognitive processing where deviations in mental workload occur. This can be accomplished through the DLR- WAT sub scales. Considering the shift from manual tasks to visual monitoring tasks, early cognitive processes concerned with acquiring and perceiving information [5] may increase workload. Workload stemming from latter cognitive processes like retrieving motor schemes for manual speed adjustment from memory is likely to decrease. By specifying possible deviations from optimal workload at each stage of cognitive processing, we can take a more informed approach towards tackling any resulting unfit workload by rethinking a certain task allocation between e.g. the train driver and ATO.

Based on these two areas, we formulated two research questions about the effects of increasing levels of automation on signaller and train driver workload. The first question answered in this research is about the magnitude and direction of overall workload deviations from an optimal level due to increasing levels of automation in railway operations. Secondly, the effect of automation on workload will be further analysed in detail on the level of the distinct stages of human information processing. These stages are represented as single scales with intermediate optima in the DLR-WAT. Thus, we formulate the following hypotheses for signallers and train drivers to answer our research questions in two separate simulator studies with professional operators.

Hypotheses for research question 1

H1.1: We expect the overall DLR-WAT scores of signallers and train drivers to deviate significantly from the assumed optimal level of workload (score of 100) in the automatic condition.

H1.2: We expect the overall DLR-WAT scores of signallers and train drivers to be lower in the automatic condition than in the manual condition.

H1.2.1: We expect the overall DLR-WAT scores of train drivers to be lower when driving with ETCS than with NTC.

H1.2.2: We expect the overall DLR-WAT scores of train drivers to be lower in the automatic ETCS condition than in the manual ETCS condition.

Hypotheses for research question 2

H2: We expect the overall DLR-WAT scores of signallers and train drivers to be lower in latter stages of information processing in the automatic condition.

Method

In order to answer the two main research questions, two separate simulator studies, one for signallers and one for train drivers, were set up and carried out. Generally, both of them featured different levels of automation, representing the development towards the highly automated scenario in rail operations described above and a baseline level for both occupational settings. To increase the validity of the studies, we followed the conclusion of Brookhuis and de Waard [23] and implemented our operational scenarios into high fidelity simulators and accepted only professional signallers and train drivers were accepted as participants in both studies. To ensure the comparability of the quantitative metrics the following measures were obtained analogously in both studies.

General measures

We assessed subjective workload through two questionnaires provided in German language, which were filled out by the participants. We used the NASA- TLX [26], as a widely used and accepted measure of workload. The NASA- TLX requires the participants to indicate their subjective workload level ranging from zero (very low) to one hundred (very high) on six subscales. The subscales are mental demand, physical demand, temporal demand, performance, effort and frustration. The mean of the numeric indications on each subscale is then computed and used as a single overall metric of workload. As a second additional measure we used the DLR-WAT questionnaire [25]. Similar to the NASA-TLX the DLR-WAT requires participants to indicate their subjective workload on a continuous scale. The DLR-WAT comprises measures on eight workload subscales. Those are information acquisition, memory retrieval and decision-making for differentiating mental workload and further physical workload, temporal workload, effort, frustration, and performance. Thus, the constructs behind the subscales are comparable to the NASA-TLX. The main difference between both questionnaires is that in the DLR-WAT there is an explicit indication of an optimal workload level in the centre of the scales and mental workload is further discriminated into three subscales according the human information processing model. In contrast to the NASA-TLX, the first six subscales from the DLR-WAT range from zero (very low) to two hundred (very high). The centre of the scale at one hundred is explicitly marked as the subjective optimal level of workload. Thus, there is an anchor of the optimal workload level in this questionnaire. The two remaining subscales differ. The subscale for frustration only ranges from one hundred (no frustration/ optimal) to two hundred (very high frustration) and the subscale for performance ranges only from zero (very low performance) to one hundred (high performance/ optimal). The idea behind the DLR-WAT is the fact that the score of one hundred always represents an optimal value. Apart from measures of workload, additional demographic data was obtained from all participants as further baseline information. The demographic data included gender, age as well as work experience as signaller or train driver.

Studies

Two separate simulator studies were carried out to compare workload scores in two distinct samples (signallers and train drivers) under different levels of automation. To answer the common research questions, both studies assessed workload via the same general measures (NASA-TLX and DLR- WAT) to ensure the comparability of the results. In the following sections short descriptions of the studies are given with regard to the participants, methods and procedure.

Study 1: Signallers

Thirteen professional German rail signallers participated in a study (N = 13) in Berlin, Germany, from which eleven were male. They were on average 25.08 years old (SD = 4.68). Nine experienced and two inexperienced signallers took part in the study. The experienced signallers (n = 9) were currently working as signallers with on average 3.93 years of work experience (SD = 4.54). The inexperienced signallers had on average 0.33 years of work experience (SD = 0.14) and were currently completing a Bachelor’s degree as part of a dual study program after completing their vocational training as a signaller. The participants received 30 € as well as the refund of their travel expenses as compensation for their participation.

The study featured two working blocks in which routine signaller tasks had to be executed with or without automatic route setting. A within- subject design with one independent variable “route setting” with two levels (manual route setting /automatic route setting) was therefore implemented in a simulated electronic interlocking. The simulated electronic interlocking is described in detail by Thomas-Friedrich, Schneider, Herholz and Grippenkoven [16]. Participants had the task of operating trains according to the timetable. The modelled infrastructure did not represent a real station, thus, the work environment was new for all participants. Before participating in the study, participants were asked to familiarize themselves with the workplace at home using a description of the experimental interlocking as well as the timetable. Upon arriving at the simulator, the experimental interlocking was explained to the participants briefly, followed by a practice period during which all train traffic had to be conducted manually by setting points and signals. Then, two experimental scenarios lasting approximately 15 minutes were completed in a randomized order. In one scenario, the *automatic route setting* system was activated, switches and points were set automatically and had to be monitored, in the other experimental scenario trains had to be *routed manually*. The timetable remained the same in practice period and experimental scenarios. After each scenario, both subjective measures of workload, DLR- WAT and NASA- TLX, were completed as dependent variables. Before leaving, the participants completed the demographic questionnaire.

Study 2: Train drivers

The second study took place in the high fidelity train simulator *RailSET* in Brunswick, Germany. A total of twenty nine professional male train drivers took part as participants. The participants were of German (n = 26), French (n = 1) and Swiss (n = 2) nationality. The mean age of the participants was 36.93 years (SD = 11.71). Participants had an average occupational experience of 12.27 years (SD = 10.65) and five of the participants had encountered ETCS in their occupation. Their main area of work was in regional service (n = 18), intercity service (n = 15) and freight operation (n = 6), but checking multiple areas of work was acceptable. The participants received 30 € as well as the refund of their travel expenses as compensation for their participation.

The study featured different driving blocks with different levels of automation and was designed as a mixed- factorial design with two independent variables “ATP” and “speed control”. The independent variable “ATP” was within-subjects and had two levels (NTC/ ETCS) while the independent variable “speed control”, which was only applicable to the parts of the track that were equipped with ETCS, was between- subjects but had two levels (manual speed control / automatic speed control) as well. The simulator is described by Stein and Naumann [27] in detail. It consists of a real cabin and panel (BR 424 Alstom/ Siemens), in which screens depicting the outside world, side windows, and the displays have been inserted. It features NTC and ETCS and was additionally equipped with an ATO system, automatically adjusting the speed according to position, speed limits, breaking curves and movement authorities. The simulated track consisted of three parts: a high speed part, followed by a rural part and leading to a final high speed part. The high speed parts were equipped with ETCS and allowed speeds up to 280 km/h and the rural part was equipped with NTC and the maximum speed was 160 km/h. Upon arrival, participants were assigned to the *manual* (n = 14) or the *automatic speed control* group (n = 15), completed the demographic data, were instructed thoroughly concerning the simulator and all systems involved and started to manually drive or monitor the ATO with the overall goal to deliver a safe and punctual service. At the end of the second part of the track equipped with NTC, participants had a scheduled stop at a station to fill in both workload questionnaires, NASA-TLX and DLR-WAT with special regard to the NTC control part of the track. After completing the last part of the track, participants filled in both workload questionnaires with special regard to the ETCS control part of the track, that was either driven manually or with ATO.

Results of study 1 and 2

To present the results in a more coherent picture, we will present results from both studies aggregated together according to our two research questions. The result section is separated in subsections for each research question. Before we report the results in light of our research questions, there are some general descriptive statistics to be reported. An aggregated overview of the data from both studies supplying descriptive means and standard deviations for both measures of overall workload, NASA-TLX and DLR- WAT, in manual and automatic conditions is presented in table 1. It becomes clear that the measures from both studies are in a format that allows comparisons between the studies, as two measures of workload (NASA- TLX and DLR- WAT scores) are supplied for signallers and train drivers alike in all of the different levels of automation. From table 1 we can see that the two measures of workload (NASA-TLX and DLR- WAT) correlated positively and significantly. This holds across both studies and across all levels of automation. For example, in the signaller study the Pearson correlation between overall NASA- TLX and DLR- WAT scores in the manual experimental block was .73. Likewise, the Pearson correlation between both measures in the train driver study in the manual ETCS condition was also .826. Since there was a strong correlation between both measures of workload and given that the DLR- WAT includes additional information concerning the optimal level of workload as well as further differentiations within the mental workload dimension, we did our inferential analyses on the basis of the DLR-WAT data.

	Study 1: Signallers				**Study 2: Train drivers**					
	Manual		**Automatic**		**NTC Manual**		**ETCS Manual**		**ETCS Automatic**	
	M	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***
NASA- TLX	39.21	14.68	32.67	17.98	42.41	15.12	35.45	14.11	29.30	9.86
DLR-WAT	103.32	15.49	82.12	20.17	95.92	18.67	74.96	26.79	65.65	26.34
Correlation	.73**		.8**		.782**		.826**		.73**	

Table 1: Descriptive statistics of gathered subjective workload measures. The shading differentiates the two separate studies. * == p < .05; ** == p < .01

Results related to research question 1

The first research question was about the magnitude and direction of workload deviations from an optimal level of workload due to increasing automation for signallers and train drivers. Therefore hypotheses H1.1, H1.2, H1.2.1 and H1.2.2 were of interest.

Turning towards the inferential part of the statistical analysis concerning H1.1 we needed to deduct whether the overall score in the DLR-WAT in the automated condition would differ significantly from an optimal level. In the DLR-WAT the optimal workload level has a score of 100. To test the hypothesis, one-sample t-tests, one for signallers and one for train drivers, were conducted. For the signaller data from study one, the overall score in the DLR- WAT differed significantly from 100 ($t(12) = -3.197$; $p < .01$), which is also depicted in figure 1. For

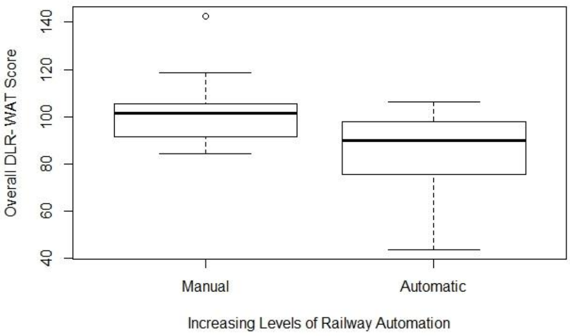

Figure 1: Overall workload scores of signallers for different levels of automation. Scores range from 0 to 200; 100 indicates an optimal workload level.

the train driver data from study two, a similar one sample t- test for the data from the highest level of automation (ETCS automatic) was done. In train drivers, the mean of the overall DLR- WAT scores also deviated significantly from 100 ($t(14) = -5.05$; $p< .01$). Indeed, figure 2 supports these effects. Thus, we confirmed hypothesis H1.1 meaning that in both studies with signallers and train drivers, the workload was significantly below the optimal level and thus underload was present while working with high levels of automation. To differentiate further between workload scores under different levels of automation, we tested H1.2, H1.2.1 and H1.2.2. H.1.2 only concerns signallers and thus only data from study one was used to test whether the overall DLR- WAT scores of signallers in study one were lower in the automatic than in the manual route setting working block. A paired t- test revealed that the overall scores differed significantly ($t(12) = 3.46$; $p < .01$). H1.2 could thus be confirmed (see again figure 1), meaning that signaller underload was significantly stronger when working with the automatic route setting than when manually setting routes.

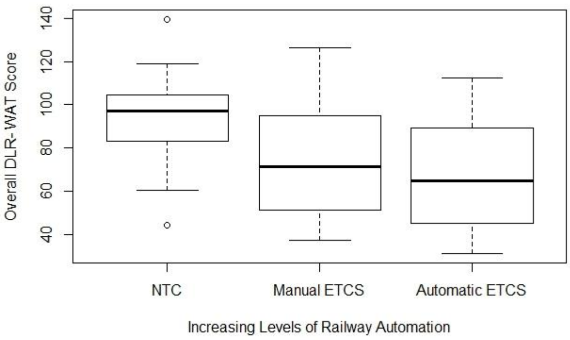

Figure 2: Overall workload scores of train drivers for different levels of automation in study 2. Scores range from 0 to 200; 100 indicates an optimal workload level.

Hypotheses H1.2.1 and H1.2.2 were concerned with the train driver workload under different levels of automation. H1.2.1 compared workload while driving with ETCS against driving with a NTC system. We decided to employ a repeated measures ANOVA with one within- subject variable comprising two levels (NTC/ ETCS), because of the mixed- factorial design of study 2. This design means that the same train drivers rated their workload in the ETCS and the NTC part of the track and therefore, we need to take that into account by employing a repeated measures model in the case. The Greenhouse- Geisser corrected effect of the within subject variable in this model proved highly significant ($F (1, 26) = 41.83$; $p < .01$). From figure 2, we could deduct the direction of this significant effect and can hence confirm H1.2.1. Overall DLR- WAT scores were significantly lower in the ETCS condition than in the NTC condition. This means that driving a train equipped with the ETCS system, which has a higher level of automation than NTC, generally led to more underload than driving a train equipped with NTC systems. Further differentiating different levels of automation within the ETCS system, we tested H1.2.2 by means of a two- sample t-test revealing no significant difference between the samples ($t (27) = 0.943$; $p = .354$). On the basis of this t- test we could not confirm H1.2.2. This means that the highest level of automation, ETCS automatic, did not result in significantly different workload ratings than the manual ETCS level of automation. Inspecting figure 2, we could observe that overall DLR-WAT scores in the automatic ETCS condition tended to be lower and therefore in the hypothesized direction, but the size of the difference does not reach statistical significance

Research question one was concerned with the magnitude and the direction of automation effects on workload. In terms of direction, it is safe to say that higher levels of automation resulted in lower workload rating. This holds in both the signaller study and the train driver study. Both operators were reporting significant underload when working within an automated setting. The magnitude of the effect was also significant in most cases. In the case of the signaller study, the magnitude of the workload difference between the two levels of automation (manual and automatic route setting) was significant. In the case of the train driver study, there were three levels of increasing automation (NTC, manual ETCS, automatic ETCS) and only a comparison of the two highest levels of automation did not reveal a significant magnitude of the effects on workload.

Results related to research question 2

The second question was about identifying sub dimensions of mental workload that strongly contribute to overall workload deviations from an intermediate optimum. Turning to the second research question, we searched further differentiation between mental workload stemming from specific stages of information processing. The first three subscales of the DLR- WAT represent mental workload stemming from information acquisition, memory retrieval and decision making. These three subscales were of interest when testing H2 separately for both studies. To analyse whether the DLR- WAT scores on the scales assessing later stages of information processing were lower in the automatic condition of signalling, a repeated measures ANOVA with two within subject variables was conducted. The first variable included the overall DLR-WAT scores on the first three sub scales representing the different stages of information processing and the second variable represented the manual and the automatic conditions. Both main effects were found to be significant, thus the overall DLR- WAT scores differed significantly between the stages of information processing ($F (1.705, 20.457) = 12.551$; $p < .01$) and for automatic versus manual work ($F (1,12) = 12.282$; $p < .01$). Inspecting figure 3, we confirmed H2 stating that lower DLR-WAT scores on the scales assessing later stages of information processing, especially memory retrieval, can be observed in signallers.

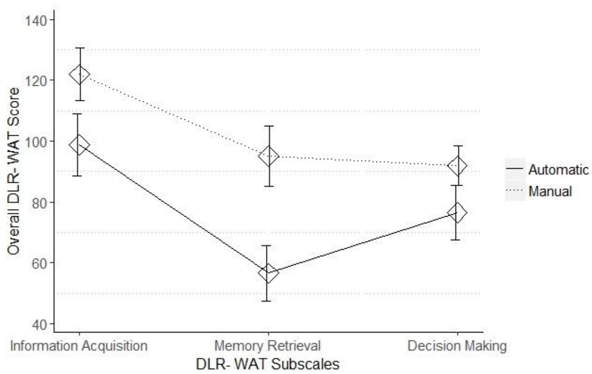

Figure 3: Differentiation of signaller mental workload stemming from different stages of information processing. Scores range from 0 to 200; 100 indicates an optimal workload level and error bars represent standard error of the mean (SE).

Figure 4 depicts the results from the train driver study on the same three sub scales, information acquisition, memory retrieval and decision- making, for the manual and automatic ETCS condition. A repeated measures ANOVA model with one within- subject variable including the overall DLR-WAT scores on the first three sub scales and one between-subject factor representing the manual and the automatic ETCS conditions was set up. It reveals a Greenhouse-

Geisser corrected effect for the within- subject variable that is just on the edge of statistical significance (F (1.58, 42.82) = 3.244; p = .059). Taking a conservative approach, we do not confirm H2 for train drivers though. Nevertheless, figure 4 shows a decrease of mental workload in the latter stages of information processing in both conditions of speed control. Especially knowledge retrieval but to a lesser extent also decision making processes contribute little to the overall mental workload, which is mainly influenced by cognitive processes related to information acquisition.

In the context of research question two, it becomes clear that latter stages of information processing generally reveal more underload or lower workload measures than the information acquisition stages. This holds for signallers and train drivers. Especially little workload is caused by memory retrieval, but also decision- making is rated as underloading. In the signaller study, this pattern is stable but workload scores are significantly lower with increasing automation. We do not observed automation effects on the three sub dimensions for the train driver.

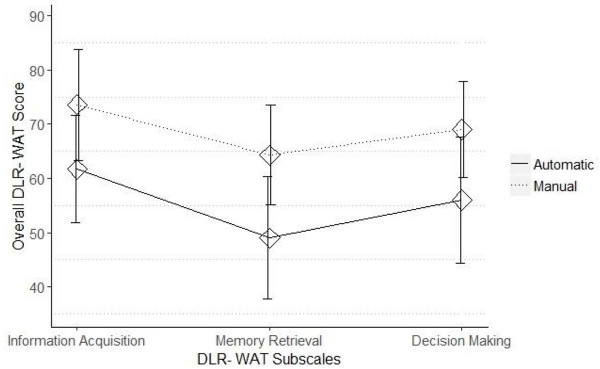

Figure 4: Differentiation of train driver mental workload stemming from different stages of information processing. Scores range from 0 to 200; 100 indicates an optimal workload level and error bars represent standard error of the mean (SE).

Discussion and conclusion

The aim of the current study is to describe the influence of increasing automation in the rail domain resulting in a growing dominance of monitoring tasks for the signaller and the train driver on their workload. Questions arise about the magnitude and direction of workload deviations from an optimal level of workload due to increasing automation. Specifically, the contribution of tasks on different stages of information processing to mental workload is of concern since visual monitoring is heavily relying on early stages of information processing as e.g. sensory perception. Identifying sub dimensions of mental workload that strongly contribute to overall workload deviations from an intermediate optimum, may help explain how performance decrements can be avoided. Focusing on how changing task environments of the signaller and train driver translate into subjective workload we can conclude on the basis of the analyses concerning the first research question that general workload measures are significantly below a subjectively perceived optimal level of workload for both signallers and train drivers in automated settings. Thus, highly automated work settings result in subjective underload of signallers and train drivers alike in the two current studies. Being able to quantify workload relative to a subjectively optimal reference level offers interesting insights and underlines the potential of the DLR- WAT. Further, hypotheses concerned with a direct comparison between manual and automatic work settings reveal lower general workload in the case of higher automation. Only one hypothesis in the train driver study fails to show this effect significantly, but the direction of the effect also points towards underload. This holds for both subjective measures obtained (NASA- TLX and DLR- WAT) as a high correlation between both measures is found throughout the analyses of the current data. Hence, a rather coherent picture emerges from these analyses. We can conclude that generally workload is significantly lower for signallers and train drivers alike while working in more automated settings than while working with lower levels of automation. Additionally, these low levels of workload are clearly perceived as underload in both occupations.

However, there are substantial differences in terms of absolute numbers between signallers and train drivers. Referring to table 1, the average DLR- WAT scores for train drivers in the manual but especially in the automated condition are of smaller size than means obtained from signallers. Hence, underload while train driving seems to be more severe than while signalling and results on hypothesis H1.2.1 suggest that changing the primary focus of attention of the train driver onto visual monitoring [9][15][11] could be a central key to explain this severe underload in train drivers driving with ATO. This finding of severe underload helps understand earlier findings in train drivers driving with ATO reporting quickly emerging fatigue [28] and poor reaction times [21][14] to unexpected events. Both are key consequences of underload described by de Waard [19] as well as Brookhuis and de Waard [23]. The same mechanism of underload but of smaller magnitude may very well be at work at the signaller workplace. Identifying these conditions of underload due to higher automation and the related consequences, which are found outside of the rail domain also [22][23][6], encourages the rethinking of the roles and tasks assigned to the train driver or signaller during the stages of stepwise implementation of automation. Alternative ideas minimizing continuous visual monitoring and proclaiming diagnosis and decision- making as central parts of the human role in rail automation have been voiced [7].

The question remains though how findings of high workload as an effect of ETCS in-cabin train driving reported by Buksh et al. [24] and Hely et al. [11] could fit into the emerging picture. Pinpointing specific stages of cognitive processing where significant deviations from optimal workload appear may help understand these seemingly contradicting findings. Findings from the study with signallers show significant decrease of mental workload stemming from latter stages of information processing like memory retrieval and decision- making in contrast to mental workload from visual perception and information acquisition. A similar effect just missing the defined criteria of significance is also seen in train drivers. There is no interaction with increasing levels of automation, but mental workload stemming from information acquisition always scores highest among the three sub domains. It could be the case that with their focus on in- cabin signalling and visual monitoring the ETCS display, Buksh et al. [24] and Hely et al. [11] studied the one subdomain of workload that is least likely to result in underload. Yet, increases in workload due to the usage of ETCS and automation cannot be replicated in the current studies. Generally, it can be concluded that mental workload stemming from latter stages of information processing, especially memory retrieval, is least pronounced in train driving and signalling. Experience with research on fatigue and ATO has shown that the novelty of ATO systems and the curiosity of participants can be possible sources of confounding activation in experimental studies possibly masking the real magnitude of underloading automation effects in experimental setups [28]. Therefore, further studies are needed at a point in time when ATO is not that new to participants anymore and occupational experience with ATO is present, in order to continue establishing a reliable indication of the real magnitude of negative automation effects of ATO in train driving. This represents one limitation of the current studies. Another limitation is the specific focus on the link between task requirements and workload without directly analysing the resulting performance parameter. This connections needs to be made in the future to evaluate a) the magnitude of possible performance decrements and b) the participants´ ability to rate their own optimal level of workload.

References

[1] Nießen, N.; Schindler, C.;Vallée, D.: Assistierter, automatischer oder autonomer Betrieb – potentiale für den Schienenverkehr, *ETR – EISENBAHNTECHNISCHE RUNDSCHAU* 4/ 2017, S. 32–37

[2] Wolbert, M.; Weiß, G.: Train Communication Network: Neue Standards für die Zugvernetzung, *ETR – EISENBAHNTECHNISCHE RUNDSCHAU* 10/ 2016, S. 49–51

[3] Brückmann, A.: Fernüberwachung von Bahntechnischen Systemen, *EI – DER EISENBAHNINGENIEUR* 4/ 2017, S. 26-27

[4] Münch, A.: Bahnkommunikationssysteme, *ETR – EISENBAHNTECHNISCHE RUNDSCHAU* 9/2016, S. 66–69

[5] Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS-PART A: SYSTEMS AND HUMANS, 30*(3), 286-297.

[6] Onnasch, L., Wickens, C. D., Li, H., & Manzey, D. (2014). Human Performance Consequences of Stages and Levels of Automation: An Integrated Meta-Analysis. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 56*(3), 476–488. doi:10.1177/0018720813501549

[7] Brandenburger, N., Naumann, A., Grippenkoven, J., & Jipp, M. (2017). Der Train Operator: Situative Fernsteuerung von automatisierten Zügen. *EI – DER EISENBAHNINGENIEUR*, 9/2017, 13-15.

[8] Pachl, J. (2017) Betriebliche Randbedingungen für autonomes Fahren auf der Schiene. *Deine Bahn*, 09/2017, 10-19.

[9] Naweed, A. (2013). Investigations into the skills of modern and traditional train driving. *Applied Ergonomics, 45*(3), 462-470. doi:10.1016/j.apergo.2013.06.006

[10] Brandenburger, N., Hörmann, H. J., Stelling, D., & Naumann, A. (2016). Tasks, skills, and competencies of future high-speed train drivers. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit.* doi:10.1177/0954409716676509

[11] Hely, M., Shardlow, T., Butt, B., Friswell, R., McIntosh, A., & Williamson, A. (2015). Effects of automatic train protection on human factors and driving behaviour. In *Proceedings 19th Triennial Congress of the IEA* (pp. 1–9).

[12] Huke, K.-P. (2017). Automatisierung und die Rolle des Fahrdienstleiters. *Deine Bahn*, 02/2017, 16-18.

[13] International Association of public Transport. (2012). *Metro Automation Facts, Figures and Trends: A global bid for automation: UITP Observatory of Automated Metros confirms sustained growth rates for the coming years*. Retrieved from www.uitp.org/metro-automation-facts-figures-and-trends

[14] Brandenburger, N., & Jipp, M. (2017). Effects of expertise for automatic train operations. *Cognition, Technology & Work, 1*(6), 59. doi:10.1007/s10111-017-0434-2

[15] Brandenburger, N., Hörmann, H. J., Stelling, D., & Naumann, A. (2016). Tasks, skills, and competencies of future high-speed train drivers. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit.* doi:10.1177/0954409716676509

[16] Thomas, B., Schneider, P., Herholz, H., & Grippenkoven, J. (2017). Measuring Rail Signaller Workload in a Highly Realistic Simulated Environment. In *Sixth International Human Factors Rail Conference* (pp. 286–295). London, UK.

[17] Young, M.S., Brookhuis, K.A., Wickens, C.D., & Hancock, P.A. (2015). State of science: mental workload in ergonomics. *Ergonomics, 58*(1), 1-17.

[18] Dobson, K. (2015). Human Factors and Ergonomics in transportation control systems. *Procedia Manufacturing*, *3*, 2913-2920.

[19] Waard, D. de. (1996). *The measurement of drivers' mental workload.* University of Groningen, Groningen, NL.

[20] da Silva, F. P. (2014). Mental Workload, Task Demand and Driving Performance: What Relation? *Procedia - Social and Behavioral Sciences, 162,* 310–319. doi:10.1016/j.sbspro.2014.12.212

[21] Spring, P., McIntosh, A., & Baysari, M. (2009). Counteracting the Negative Effects of High Levels of Train Automation on Driver Vigilance. *HFESA Conference Proceedings, 45,* 93-101.

[22] Endsley, M., & Kaber, D. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462–492.

[23] Brookhuis, K. A., & Waard, D. de. (2010). Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention, 42*(3), 898–903. doi:10.1016/j.aap.2009.06.001

[24] Buksh, A., Sharples, S., Wilson, J. R., Morrisroe, G. & Ryan, B. (2013) Train automation and control technology-ERTMS from users' perspectives. In Anderson, M. (Ed.) (2013) *Contemporary Ergonomics and Human Factors 2013.* UK, London: Taylor and Francis

[25] Grippenkoven, J., Rodd, J., & Brandenburger, N. (in press). DLR – WAT: Ein Instrument zur Erhebung der Beanspruchung in hochautomatisierten Mensch – Maschine Systemen. *AAET 2018*.

[26] Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. Advances in Psychology, 52, 139-183.

[27] Stein, J. & Naumann, A. (2016). Monotony, fatigue and microsleeps—train drivers` daily routine: a simulator study. In Milius, B. & Naumann, A. (eds): *Proceedings of the 2nd workshop on Rail Human Factors*, Brunswick, Germany, pp 96–102.

[28] Brandenburger, N., Wittkowski, M., & Naumann, A. (2017). Countering Train Driver Fatigue in Automatic Train Operation. In *Sixth International Human Factors Rail Conference* (pp. 57-65). London, UK.