Predicting Take-Over Times of Truck Drivers in Conditional Autonomous Driving

Alexander Lotz(✉) and Sarah Weissenberger

Daimler AG, Daimler Trucks Advanced Engineering,
HPC T332, 70546 Stuttgart, Germany
{rene_alexander.lotz, sarah.weissenberger}@daimler.com

Abstract. Conditional autonomous driving requires the description of sufficient time reserves for drivers in take-over situations. The definition of this time reserve has not been addressed for the truck context thus far. Through the observation of physiological measures, the possibility of estimating reaction times is considered. Driver data is collected with a remote eye-tracker and body posture camera. Empirical data from a simulator study is utilized to train and compare four machine learning algorithms and generate driver features. The estimation of take-over times is defined as a classification problem with four reaction time classes, leading to a misclassification rate of a linear support vector machine (SVM) of 38.7%. Utility of driver features for reaction time estimation are discussed.

Keywords: Conditional autonomous driving · Take-over prediction
Human factors · Machine learning · Control transition

1 Introduction

One of the prime interests of recent automotive studies is how attention, communication and responsibilities are allocated, when humans and machines share control of a vehicle. While this surge prevails predominantly for passenger cars, the truck context has not been regarded as intensely. Currently, the definition according to SAE J3016 [1] classifies six stages of autonomy for vehicles. Conditional autonomous driving (CAD), which classifies as Level 3, and higher levels allow drivers to disengage from the driving task and focus their attention on non-driving-related tasks (NDRT). Thereby, possibly, the most critical stage is Level 3, in which drivers are required to act as a fallback if system limits are met and accordingly only receive restricted time to take back control. Following the definition of CAD, drivers are no longer required to constantly supervise the vehicle when the automation function is activated. Adapting the responsibility of constant monitoring of the truck would revolutionize logistics and leave drivers to focus on different aspects of their work (e.g. paper work) and other NDRT. At take-over, when the automation function can no longer handle the driving situation, drivers are expected to perceive, address and solve given situations in limited time. Due to the switch of control between driver and vehicle, human factors such as the ironies of automation [2], have to be taken into account. Generally, the truck context offers a more confined set of use cases with less variation in the type of driving

N. Stanton (Ed.): AHFE 2018, AISC 786, pp. 329–338, 2019.
https://doi.org/10.1007/978-3-319-93885-1_30

situations that have to be managed by the advanced driving assistant system providing autonomous driving. Long distance hauls by heavy duty trucks on motorways comprise 21% of freight traffic in Germany and professional drivers' duties [3]. The implementation of a Level 3 system could entail the possibility for drivers to engage in NDRT, improve average traffic flow rates and cause a change in legislation, currently restricting operating times for drivers in some countries, e.g. Germany [4]. As take-over situations vary and driver attention to NDRT is not an inter- or intra-individual constant, the system has to allow for all drivers to transition back to the driving task. Time reserves for an adequate transition have to either be defined globally, with a fixed time reserve sufficient for all drivers, or individually. For individual time reserve definition, ideally, reactions by the driver would be predictable through objective measures, allowing the system to calculate take-over times. This definition of time reserves is far more complex, as all or the most influential factors need monitoring and consideration. Vogelpohl et al. [5] identify four groups of influencing factors on reaction times that will be regarded in this study. (1) *environmental* parameters (e.g. traffic density, weather, complexity of take-over situation), (2) *human machine interaction* (HMI) parameters (e.g. warning modalities, interfaces), (3) *driver* parameters (e.g. visual attention, demographics, individual driving style) and (4) *vehicle* parameters (e.g. field of vision, advanced driver assistance systems).

1.1 Take-Over Times

A variety of take-over times are relevant in the take-over process and defined by Damböck [6] at the occurrence of a "Request to Intervene" (RtI) in CAD. These times include the time until visual attention is shifted back to road (time to eyes on road; TTEoR), time until hands are placed on the steering wheel (time to hands on steering-wheel; TTHoS) and time until a first reaction is given at the steering wheel or accelerator pedals (time to first reaction; TTFR). The measurement of said times allow the evaluation of the take-over process, consisting of physiological attention allocation and cognitive processing [7].

1.2 Prediction of Driver Reactions

Although driver assistance systems should reduce stress and accidents, an ill chosen time reserve at take-over can paradoxically cause the opposite [8]. It is paramount to guarantee this adequate time reserve. Wulf et al. [9] identify the cognitive and motoric availability as essential for take-over readiness. Additionally, [10, 11] show that situation awareness [12] is decisive for take-over times and the quality. Therefore, the implementation of a driver observation could offer relevant data to guarantee sufficient take-over times, as one of the four influencing factors on reaction times. It is the goal of this study to investigate the feasibility of training machine learning algorithms to predict take-over times.

In order to predict take-over times, this study focuses on the recording of behavioral driving data during CAD in heavy duty trucks and investigates the features generated from these recordings. In the driving context of CAD, Braunagel et al. [13] show that machine learning can be applied to identify four different NDRT, based on eye-tracking, with an accuracy of 82%. Apart from eye-tracking, Ohn-Bar et al. [14] additionally investigates hand and head patterns to detect where drivers are interacting. In an extension of their work, Braunagel et al. [15] present an architecture classifying the driver readiness and considering traffic situations for CAD, reaching an overall accuracy of 79%. None of these known architectures have been tested for trucks or the prediction of reaction times.

2 Estimation Concept for Reaction Times

A concept to estimate reaction times at RtI during CAD with NDRT is be presented. All necessary ground truth data for such a concept was collected during a simulator study in the Mercedes-Benz Moving-Base-Simulator with professional truck drivers (n = 88). The experimental setup was designed to generate multiple take-overs per participant, producing data for our estimation approach. When considering the abovementioned four groups of influencing factors on reaction times by [5], the experimental setup was configured to have differentiating values for *environmental* and *driver* parameters, while keeping *HMI* and *vehicle* parameters constant. During CAD, participants were encouraged to interact with two NDRT, to allow for realistic distraction. The available NDRT were either a geography game or a video.

Eye-tracking is a common method for collection of behavioral data, especially important during the dynamic driving task [16]. In order to accumulate data distraction free, i.e. non-invasive, a remote eye-tracker (Smart Eye Pro) collected gazes and head position. Additional interaction by the driver with the environment was collected through a fixed tablet (non-nomadic device) on which NDRT are presented. A Microsoft Kinect gathered RGB-D pictures. These pictures were utilized to calculate participants' body posture during their CAD drive. The extraction of body posture is an adaptation of [17], in which a pre-trained convolutional neural network is retrained for the truck context. The second group of influencing factors, *environmental* parameters [7] are taken into account through the simulator. Examples of these parameters are the take-over situation and the time to collision (TTC), i.e. the lead time to obstacles at RtI.

2.1 Online and Offline Datasets

Two different datasets are generated from the eye-tracking, body posture and simulator data. When considering 755 take-over procedures, each trial is divided into a time series before the RtI in which the driver is not required to steer the vehicle and a time series when manual control is necessary by the driver after the RtI. After an RtI is issued, certain additional features can be calculated that depend on the occurrence of an RtI and the time series in which the driver maintains manual control. This includes all the reaction times such as time to eyes on road (TTEoR) and time to hands on steering wheel (TTHoS) apart from TTFR, which is the desired estimation output. These features may hold high contents of information that generate higher estimation results due to temporal interrelation. However, such features are not applicable for the online estimation of reaction times, as they can only be considered after the presentation of an RtI. Therefore, two separate datasets are configured:

- *Online dataset*: with features computable at any instance during CAD
- *Offline dataset*: all computable features for each trial.

2.2 Feature Generation

To allow the estimation of TTFR, features need to be generated holding preferably high contents of information about driver and/or environment. Prior to training machine learning algorithms with manually configured features, a selection according to correlation is performed. All features are either static singular values or a time series for each of the 755 take-over situations (trials). The time shortly before an RtI probably holds the most information with regard to the take-over capability of a driver and will be considered. To bypass the possibility of manually defining a time interval that may not hold any relevant information, four different time intervals are considered; the complete CAD trial duration, 10 s, five seconds and two seconds before an RtI. Singular values, such as average, standard deviation, minimum and maximum of the abovementioned time intervals are calculated. This methodology is applied for the feature generation of eye-tracking, body posture and simulator data.

The feature generation of eye-tracking and simulator data yields a total of 238. In addition, the calculation of features from extracted body posture of 12 body joints yields 135 further features.

2.3 Feature Evaluation and Selection

Reaction time TTFR is subdivided into four time classes for the application of supervised machine learning classification algorithms. Figure 1 displays a graphic representation of a frequency distribution of all 755 TTFR. The boundaries of the four

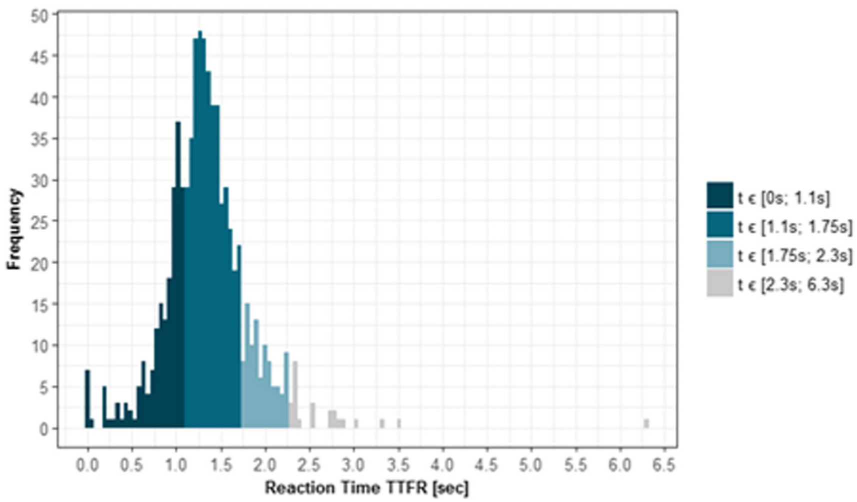

Fig. 1. Frequency distribution of time to first reaction (TTFR). Reaction times are categorized according to the definition found in Table 1 by application of the clustering algorithm k-means (k = 4).

TTFR classes are calculated by application of the k-means algorithm. The classes are partitioned as presented in Table 1.

Table 1. The dataset of all 755 take-overs are subdivided into four classes through a k-means algorithm.

Class	Definition	Number of Take-Overs
1	t ϵ [0 s; 1.1 s]	212
2	t ϵ [1.1 s; 1.75 s]	430
3	t ϵ [1.75 s; 2.3 s]	91
4	t ϵ [2.3 s; 6.3 s]	22

A filter approach is applied for the evaluation of all previously generated features. A multivariate analysis of variances (MANOVA) is implemented to identify those features that show the highest correlation between input and output. A total of 30 features with the highest congruence for TTFR are selected for the application of the four different machine learning algorithms. Table 2 holds these features for the *online dataset*. Notably, by analysis of the offline dataset only one feature is replaced in this list. Time to hands on steering-wheel (TTHoS) holds the most information for the estimation of TTFR in the offline dataset and is ranked first.

Table 2. The 30 most important features selected from the Online dataset through application of MANOVA.

Rank	Feature	Rank	Feature
1	Trial Number (# of take-overs)	16	Number of Blinks 10 s before RtI
2	Time to Collision	17	Avg. Head Heading
3	Blink Duration	18	Max Gaze Direction Change 5 s before RtI
4	Avg. Head Heading 2 s before RtI	19	Head Roll 0.5 s before RtI
5	Take-over Situation type	20	Avg. Head Roll
6	Head Roll Average 2 s before RtI	21	Number of Blinks 5 s before RtI
7	Min Head Pos. Change X-Y-Z	22	Blink Duration 5 s before RtI
8	Head Pos. Change X 0.5 s before RtI	23	Max Head Pitch 2 s before RtI
9	Head Pos. Change Y 3 s before RtI	24	Head Pitch 1 s before RtI
10	Min Head Pos. Change X-Y-Z 10 s before RtI	25	Min Head Pos. X-Y-Z 2 s before RtI
11	Pupil Diameter 0.5 s before RtI	26	Min Head Pos. Change 2 s before RtI
12	Gaze Pitch St. Dev. 5 s before RtI	27	Head Pitch 1.5 s before RtI
13	Sum of tablet taps 5 s before RtI	28	Min Head Pos. X-Y-Z 10 s before RtI
14	Avg. Head Pos. X-Y-Z 2 s before RtI	29	Max Head Pitch 5 s before RtI
15	Gaze Pitch St. Dev Total	30	Avg. Gaze Direction X

2.4 Machine Learning Algorithm

Four different machine learning methods will be analyzed for this applied problem. For the estimation of TTFR, which is defined as a classification problem, the following machine learning algorithms are tested:

- k-Nearest Neighbors algorithm
- Bayes classifier
- Support Vector Machine
- MLP Neural Network

In terms of a machine learning problem, the dataset collected through this simulator study with 755 instances is rather small. Complex algorithms with a high share of hyper parameters are not suited for this problem as there is no abundance of data. An interaction between many of the abovementioned different features is expected for the estimation of the output variable, i.e. TTFR. Therefore, a trade-off between the complexity of the algorithm and the size of the dataset has to be accepted. Two simple and two more complex machine learning algorithms are chosen.

All tested algorithms are validated by means of a stratified n-fold cross-validation. As proposed by [18], a 10-fold cross-validation is chosen. The complete dataset is divided into a training and test dataset, with a distribution of 80% to 20% respectively.

3 Results

As a quality measure for the classification by all machine learning algorithms, misclassification rate is chosen. The misclassification rate hereby describes the proportion of misclassified observations that were predicted into one of the four TTFR classes, see Table 1 and Fig. 1.

As the estimation of the online dataset bears the most relevance for the CAD context, detailed results are presented. A comparison between additional information from body postures and to the offline dataset are produced.

3.1 Online Estimation

In the following, all four machine learning algorithms will be compared by means of misclassification rates for the online dataset. Especially the third and fourth class are not represented evenly in the data (C3: 12%; C4: 3%), see Table 1. While these times are more seldom due to the Gaussian distribution of reaction times for this study, especially in the context of take-over times these instances are the most relevant as they represent critical take-overs due to slow reaction times. To compensate for the unbalanced classes of the dataset, a cost matrix is introduced. This facilitates weighted penalties for misclassifications.

k-Nearest Neighbours: For the definition of the optimal k-NN the hyper parameters k and the distance function are iterated. The Mahalanobis distance function exceeded the results of the Euclidean distance function. Best results were received for $k = 8$, leading to a misclassification rate of 40.8%.

Bayes classifier: Again, the Mahalanobis distance function provided the best results. A feature scaling did not improve results. Minimal misclassification was achieved at 39.6%.

Support Vector Machine: Three different SVM types are tested, i.e. linear, Gaussian and polynomial. Hereby, the linear and Gaussian SVM achieved a misclassification rate of 38.7%, while the polynomial SVM performed 39.2%.

Neural Network: The optimal misclassification rate for the neural network of 40.0% was achieved with one layer and five nodes.

Overall, the best results are obtained for the linear SVM with a misclassification rate of 38.7%. Table 3 displays a comparison of the classification results of the tested algorithms.

Table 3. Comparison of the best classification results of four tested machine learning algorithms on online dataset

Algorithm	Misclassification rate [%]
k-NN	40.8
Bayes	39.6
SVM	38.7
NN	40.0

3.2 Offline Estimation

The following estimation on the offline dataset is obtained with TTHoS ranked first in the list of 30 most relevant features, see Table 2. Similarly to Sect. 3.1, the SVM obtains the lowest misclassification rate at 22.5% (Table 4).

Table 4. Comparison of the best classification results of four tested machine learning algorithms on offline dataset

Algorithm	Misclassification rate [%]
k-NN	38.3
Bayes	32.6
SVM	22.5
NN	26.4

3.3 Online Estimation with Body Posture

As the SVM delivered the lowest misclassification rate for online estimation without body posture features, no comparison of machine learning algorithms will be provided. Similar methodology as in Sect. 3.1 is applied. After application of a cost matrix the linear SVM misclassification rate presented a value of 37.7%. The consideration of the calculated 30 features shows that 18 new features (body posture features) are replaced in the list presented in Table 2. The highest of these new features ranking in fourth position and representing the maximal displacement of the right knee in the 10 s before RtI.

4 Discussion

Comparing the results of the online and offline classifications without body posture, see Sects. 3.1 and 3.2, a large difference of 16% between both linear SVM exists. As first ranked TTHoS is the only feature presented in the offline dataset not being available during online estimation, it seems to hold a lot of relevant information. This is not surprising, as TTFR partially depends on the motoric hand availability for adequate reaction times. Overall 76% of TTFR are induced due to a motoric reaction at the steering wheel in the data, while only 24% of first reactions were cause by braking or accelerating. While the discrepancy of the algorithm comparison for the online dataset were within 2.1%, this can be attributed to the absence of highly relevant features. In other words, the features were similarly meaningful. TTHoS however does not hold 16% more information in the offline dataset, the algorithms SVM and NN can attribute higher relevance to certain features by redistributing weights.

Based on the features calculated and due to the fact that the reaction time classes were unequally distributed, the quality of the results obtained is not sufficient to safely and reliably estimate reaction times of truck drivers. By placing a new observed reaction time into the second class, 43% of these instances the estimation would misclassify, (430 trials were in this class, see Table 1). Both the online and offline classification achieved a lower misclassification rate, showing that some information is drawn from the features calculated. There are multiple reasons why only a small reduction in misclassifications was achieved. Firstly, the quality of the data may not be sufficient enough. Either the static features calculated by hand or the variance in the data can cause a reduction of the signal quality. Secondly, the problem may be too complex for an analysis by the evaluation methods of this study. Thirdly, a major problem when addressing such a problem is the generation of data. Production of sufficient amounts of data needs time. Especially for problems in which the most relevant data, i.e. slow reaction times, is sparse, calling for extensive data collection. Lastly, human behavior may not be classifiable and features bear no consisted information on reaction times.

To the best of the authors' knowledge, no studies for CAD are published that required nine or more take-overs within 30 min. The data shows, that the trial number was categorized as a feature with high correlation to TTFR. When looking at this variable separately, learning behavior is identified. Participants might have therefore been cautious after the first couple of take-overs resulting in such quick and dense reaction times (TTFR) as seen in Fig. 1.

The additional consideration of body postures, represented by the localization of 12 body joints showed only a minimal improvement of 1% in the misclassification rate of a trained linear SVM for the online dataset. Although more than half the elements in the features list were represented by body posture values, this new remote observation of physiological signals showed minimal information gain. This could mean that body postures before a take-over hold no relevant information. Another possible explanation would be that the experimental setup did not encourage a large variety of body postures. According to *Fitts' Law* [19] movement is a function of target distance and target size. Consequently, the motoric variation seems to have been minimal in this experimental setup and did not cause for large time discrepancies. Finally, the body joint estimation method could entail flaws and supply false coordinates leading to lower signal quality.

Overall the results of this study were surprising, as times to first reactions were far below those reported in current CAD publications. Neither of the independent variables, CAD duration and NDRT, presented significant differences and 99,7% (3σ-interval) of all take-overs were completed within 2.82 s. Our work shows that although information can be drawn from static features from eye-tracking and body posture analysis, no clear correlation between multiple features and reaction times can be drawn. The improvement of measurement technology and feature calculation could reduce the misclassification rate further; however, it is unlikely that full classification can be achieve. Variance in drivers' actions seems too high for a current mathematical prediction method.

4.1 Future Work

This publication is limited to the analysis of take-over time through the implementation of classification algorithms. In future work, the problem could be considered as a regression problem, possibly leading to a better estimation. Secondly, a new classification problem with balance reaction time classes and probabilities will be computed. This, however, presents the drawback that classes with higher standard deviation from the expected Gaussian value will become large. Thirdly, non-static features will be considered for analysis. Distinct correlations between features that were not defined in the generated feature list of this work could possibly provide a lower misclassification rate.

References

1. SAE On-Road Automated Vehicle Standards Committee. Information Report J3016: Taxonomy and Definitions (2014). https://www.sae.org/misc/pdfs/automated_driving.pdf
2. Bainbridge, L.: Ironies of automation. Automatica **19**, 775–779 (1983)
3. Statistisches Bundesamt. Verkehr auf einen Blick, Wiesbaden, Germany (2013)
4. VO (EG) Nr. 561/2006 of the European Parliament (2006)
5. Vogelpohl, T., Vollrath, M., Kuehn, M.: Uebergabe von hochautomatisiertem Fahren zu manueller Steuerung. Forschungsbericht **47** (2017)
6. Damböck, D.: Automationseffekte im Fahrzeug – von der Reaktion zur Übernahme. München, Technische Universität München, Dissertation (2013)
7. Petermeijer, S.M., Hornberger, P., Ganotis, I., de Winter, J.C.F, Bengler, K.J.: The Design of vibrotactile seat for conveying take-over requests in automated driving. In: Stanton, N.A. (ed.) Advances in Human Aspects of Transportation. Advances in Intelligent Systems and Computing, vol. 597, pp. 618–630. Springer, Heidelberg (2018)
8. Regan, M.A., Lee, J.D., Young, K.: Driver Distraction: Theory, Effects, and Mitigation, pp. 85–105. CRC Press, Boca Raton (2008)

9. Wulf, F., Zeeb, K.: Effects of human-machine interaction mechanisms on situation awareness in partly automated driving. In: 16th International IEEE Annual Conference on Intelligent Transportation Systems, pp. 2012–2019 (2013)
10. Rudin-Brown, C.M., Parker, H.A.: Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. Transport. Res. F Traffic Psychol. Behav. **7,** 59 (2004)
11. Radlmayr, J., Gold, C., Lorenz, L., Farid, M., Bengler, K.: How traffic situations and non-driving related tasks affect the take-over quality in highly automated driving. In: Proceedings of the Human Factors and Ergonomics Society 58th Annual Meeting, pp. 2063–2067 (2014)
12. Endsley, M.R.: Toward a general theory of situation awareness in dynamic systems. Hum. Factors **37**(1), 32–36 (1995)
13. Braunagel, C., Rosenstiel, W., Kasneci, E.: Ready for take-over? A new driver assistance system for an automated classification of driver take-over readiness. In: IEEE Intelligent Transportation Systems, pp. 10–22 (2017)
14. Ohn-Bar, E., Martin, S., Tawari, A., Trivedi, M.M.: Head, eye, and hand patterns for driver activity recognition. In: 22nd International Conference on Pattern Recognition, pp. 660–665 (2014)
15. Braunagel, C., Rosenstiel, W., Kasneci, E.: Online Recognition of Driver-Activity based on Visual Scanpath Classification. In: IEEE Intelligent Transportaion Systems, pp. 23–36 (2017)
16. Gold, C., Damböck, D., Bengler, K., Lorenz, L.: Partially automated driving as a fallback level of high automation. In: Tagung Fahrerassistenz, vol. 6 (2013)
17. Insafutdino, E., Pishchulin, L, Andres, B., Andriluka, M., Schiele, B.: Deeper cut: a deeper, stronger, and faster multi-person pose estimation model. In: European Conference on Computer Vision, pp. 34–50 (2017)
18. Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. Int. Joint Conf. Artif. Intell. **2**, 1137–1143 (1995)
19. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. J. Exp. Psychol. **47**(6), 381–391 (1954)