sensors

MDPI

*Article*

# The Identification of Non-Driving Activities with Associated Implication on the Take-Over Process

**Lichao Yang [1], Mahdi Babayi Semiromi [1], Yang Xing [1], Chen Lv [2], James Brighton [1] and Yifan Zhao [1,*]**

[1] School of Aerospace, Transport and Manufacturing, Cranfield University, Bedford MK43 0AL, UK; lichao.yang@cranfield.ac.uk (L.Y.); m.babayi-semiromi@cranfield.ac.uk (M.B.S.); Yang.X@cranfield.ac.uk (Y.X.); j.l.brighton@cranfield.ac.uk (J.B.)

[2] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore; lyuchen@ntu.edu.sg

* Correspondence: yifan.zhao@cranfield.ac.uk

**Abstract:** In conditionally automated driving, the engagement of non-driving activities (NDAs) can be regarded as the main factor that affects the driver's take-over performance, the investigation of which is of great importance to the design of an intelligent human–machine interface for a safe and smooth control transition. This paper introduces a 3D convolutional neural network-based system to recognize six types of driver behaviour (four types of NDAs and two types of driving activities) through two video feeds based on head and hand movement. Based on the interaction of driver and object, the selected NDAs are divided into active mode and passive mode. The proposed recognition system achieves 85.87% accuracy for the classification of six activities. The impact of NDAs on the perspective of the driver's situation awareness and take-over quality in terms of both activity type and interaction mode is further investigated. The results show that at a similar level of achieved maximum lateral error, the engagement of NDAs demands more time for drivers to accomplish the control transition, especially for the active mode NDAs engagement, which is more mentally demanding and reduces drivers' sensitiveness to the driving situation change. Moreover, the haptic feedback torque from the steering wheel could help to reduce the time of the transition process, which can be regarded as a productive assistance system for the take-over process.

**Keywords:** non-driving related activity (NDRA) classification; level 3 automation; 3D CNN; take-over transition; situation awareness

check for updates

**Citation:** Yang, L.; Babayi Semiromi, M.; Xing, Y.; Lv, C.; Brighton, J.; Zhao, Y. The Identification of Non-Driving Activities with Associated Implication on the Take-Over Process. *Sensors* **2022**, *22*, 42. https://doi.org/10.3390/s22010042

Academic Editor: Weizhi Meng

Received: 25 November 2021
Accepted: 20 December 2021
Published: 22 December 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Conditional automation systems (level 3), defined by the SAE (J3016) Automation Levels [1], releases the driver's eyes and hands from monitoring the environment and controlling the vehicle. Such systems can perform some non-driving activities (NDAs) during automated driving, however, they would have to intervene in the control of the vehicle when requested. Even though lots of level 2 and 3 automation systems have been commercialised in the automotive industry, the immature design and the excessive trust of the driver still cause accidents, even costing lives. Two Tesla fatalities occurred in Williston, Florida, USA, 2016 and Mountain View, California, USA, 2018. In both fatalities, the Autopilot systems were engaged, and the drivers were performing some NDAs before and when the accident happened (watching movies and playing games). Neither the Autopilot nor the driver noticed the hazard ahead and took action to avoid the accident, even though there was sufficient time and distance to react to prevent the crash [2,3]. Both fatalities could have been avoided if there was a driver monitoring and alert system to prevent the prolonged disengagement of the dynamic driving task. Since the driver's situation awareness could be reduced and their mental demand could be increased by the NDAs engagement [4–6], automatically recognising the driver's NDAs engagement and further understanding its impact on the take-over performance is of great importance to design an intelligent human–machine interface (HMI) for a safe and smooth take-over process.

In terms of NDA recognition, the most similar research is secondary task detection, which is an important part of driver behaviour monitoring for conventional vehicle safety. It aims to evaluate the degree of distraction while the driver is driving. The related methods can be roughly categorized into three groups, including vehicle maneuver-based methods, driver gesture modelling methods, and information fusion methods. The vehicle maneuver-based methods measure the vehicle state features, e.g., speed, longitudinal acceleration, lateral acceleration, pedal position, etc. Such methods provide high accuracy in terms of the engagement detection. However, the performance in terms of identifying a specific task is relatively poor [7,8]. The gesture modelling methods directly model the driver's body pose to classify the engaged task [9,10]. The information fusion methods extract both features from the driver's behaviour (e.g., head or gaze movement) and the vehicle state characteristics to achieve robust task detection [11,12]. Compared with the secondary task detection, the NDA recognition aims to identify the specific activity the driver is engaging in while the vehicle is in the automated driving mode, which could affect the take-over performance. The existing common approaches used in secondary task detection suffer from the limitation of action classification. For instance, during NDA engagement, the driver is free from controlling the vehicle. The maneuver-based methods are therefore not able to capture the driver's distraction. It is also a challenge to refine activity recognition using a single gesture or gaze modelling method due to the diversity and uncertainty of NDA engagement [13].

Previous studies of drivers' take-over performance claimed that a sufficient take-over interval for the driver should be between 5 and 8 s [6,14]. The factors that influence the performance include the driver's state, such as gender, age, and driving experience [15,16], the complexity of the driving scenario [17–19], the modality of the take-over request [14,20,21], and the NDAs that the driver engages in [14,22]. In recent years, the effect of NDAs on the take-over transition process has been broadly investigated. Yooh et al. [14] investigated drivers' take-over performance for three types of NDA, namely phone conversation, smartphone interaction, and video watching tasks, while Zeeb et al. [23] examined the impact of writing an email, reading news, and watching video clips. Results from both studies suggested that NDA engagement can significantly affect take-over quality based on the statistical analysis. One of the limitations of existing studies [24,25] is that NDAs were investigated specifically and independently. When considering a new NDA, such a system needs to conduct the evaluation process again, which limits the extendibility of the driver monitoring or take-over assistance system. There is a lack of systematic methods to group NDAs which could have a similar level of impact on the take-over performance for enhanced scalability. On the other hand, the existing literature concerning the impact of NDAs is normally from the perspective of the driver's workload [6,26,27]. Situation awareness before take-over is also considered as a crucial factor of safe take-over transition but has not been discussed in association with NDAs [28]. There is a knowledge gap concerning the implication of situation awareness on the take-over process.

It has been stated in the existing literature that the type of NDA that drivers engage in affects their take-over. For instance, compare to auditory related activities, visual related activities lead to a longer reaction time [29]. However, the existing research only focused on a specific visual related task and an auditory related task. There is no further discussion on more different visual related or auditory related activities. Following the survey undertaken by Sivak and Schoettle [30], the identified common NDAs are reading, texting, working, watching movies, and playing games. Since all these common activities are visual related, in this study, we focus on this kind of NDA specifically. The four selected types of NDA include playing games, answering questionnaires, watching videos, and reading news. The device that the driver used to engage all these NDAs is a tablet. Based on the type of interaction between the driver and tablet, the selected NDAs are divided into two groups: active interaction mode and passive interaction mode. Playing games and answering questionnaires are considered as the active interaction mode since the driver and the object respond to each other's action over time during the engagement.

Under the passive interaction mode, e.g., reading news or watching videos, the driver only receives information passively. For simplicity, the active NDAs and passive NDAs will be used to represent the NDAs in the active interaction mode and passive interaction mode, respectively. This study hypothesizes that the workload and demanded attention are different between these two modes, which will lead to different take-over performance.

This paper proposes a two-feed, computer vision-based framework for NDAs/DAs recognition with 3D convolutional neural network (CNN). The driver's behaviour, including head movement and hand movement, is considered as the input of the CNN framework. Then, the implication of the recognized NDAs in both interaction modes on the take-over performance was investigated. Additionally, based on the captured head movement video, the driver's road-checking behaviour has been extracted, which is considered a factor that reflects the driver's situation awareness. If more road-checking behaviour has been performed during the engagement of NDAs, the driver will have more awareness of the surrounding environment. The motivation for performing such behaviour associated with each NDA has been inferred for further understanding the attention demand under different NDAs. The haptic feedback torque was implemented in the steering wheel to support the driver in the control transition. The haptic feedback assistance during the take-over process has been broadly investigated regarding the design of the HMI [31–33], especially the steering wheel implemented system [34–36]. This study also evaluated its impact and effectiveness in the take-over process.

The paper is organized as below. The NDA detection and recognition system, the experiment design, and the vehicle setting are introduced in Section 2. In Section 3, the performance of the 3D CNN model used in this study is evaluated. Furthermore, the driver's road-checking behaviour and their take-over performance of each NDA are presented and analysed from both the perspective of the group and individual. The discussion and conclusions are further provided in Section 4.

## 2. Methodology

### *2.1. NDA Detection and Recognition System*

The driver's activities inside the cabin can be divided into 2 groups, which are driving-related activities and non-driving related activities. Gaze estimation is one of the most commonly used methods for the detection of driving-related activities, such as road checking, wing mirror checking, and rear-view mirror checking [37–39]. This method focuses on modelling through the driver's facial features. It can work effectively in detecting the driver's road-checking behaviour and identifying some activities, e.g., centre console checking and dashboard checking. However, for the application on NDA recognition, such a method can only identify which object (e.g., phone or tablet) that the driver is gazing at but cannot recognise the specific task [40], such as whether the driver is watching videos or playing games with this object, which could lead to different take-over performances [6,23]. Yang et al. [41] proposed a two-stream CNN model for NDA recognition based on the driver's hand movement. The spatial stream demonstrates high performance on the classification of objects (phone and tablet). The temporal stream uses a stack of optical flow frames which represent the hand movement between 2 RGB frames. Xing et al. [42] further extracted the driver's body from the RGB frames and used the segmented body frames as the input of the CNN model to recognize the driver's behaviors. Both methods are based on the 2D CNN model, in which the convolutions only consider the features from the spatial dimension of the frames [43]. The features of hand movement behaviour in a time duration provided in the optical flow stack can only be processed as multiple channels in the spatial dimension. It lacks a direct representation of the motion information in the temporal dimension.

To address the above problem, this paper introduces a 2-feed 3D CNN-based NDA recognition framework, the flowchart of which is shown in Figure 1. This framework is used to recognize 4 types of NDAs and 2 types of driving activities (DAs). The details of these activities will be introduced in the section below. The front camera captures

the driver's head movement, the image of which is cropped based on the face location automatically. A stack of head movement frames within a certain time window, denoted as $d$, are inputted to a 3D CNN model to detect whether the driver is performing NDAs or DAs. The rear camera focuses on the driver hand movement. The frame is fixedly cropped since the hand movement is limited in the vehicle cabin. A stack of hand movement frames within the same time window is imported into 2 separate models to further identify the specific NDA or DA. The final prediction result is obtained from the concatenated results of the NDA and DA classification.

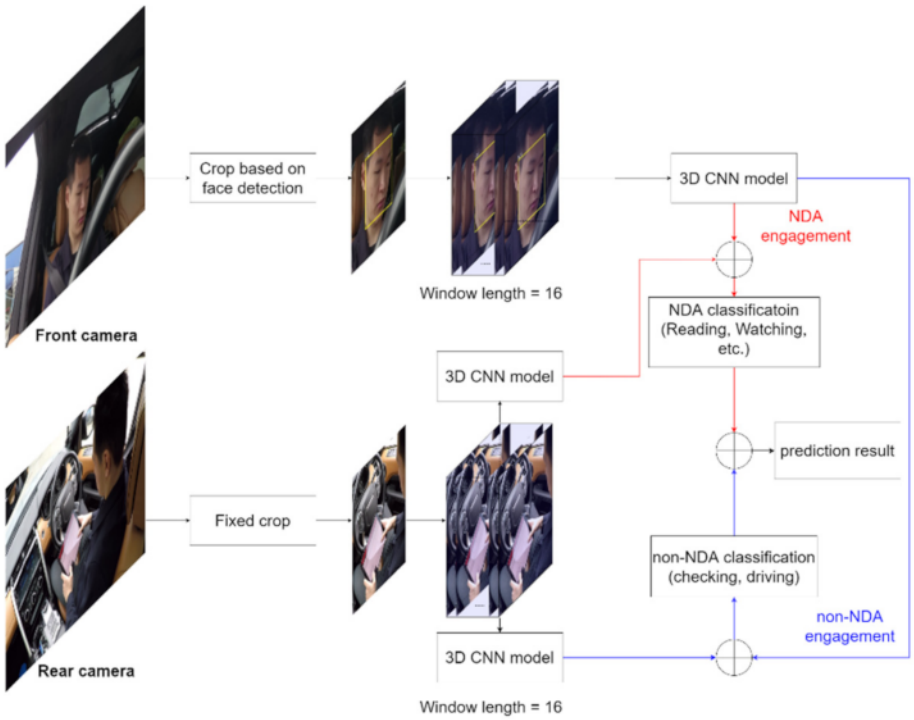


**Figure 1.** The proposed 2-feed NDA recognition framework.

The architecture of the network is based on 3D ResNet-18, whose capability in terms of video recognition has been proven in [44,45]. The structure is illustrated in Figure 2. The size of the input frame stack and the feature map is notated as $c \times d \times h \times w$, where $c$ is the number of the channels, $d$ is the depth of the input (the number of frames in the time window for this case), $h$ is the height of the frame, and $w$ is the width of the frame. The convolutional kernel size is denoted as $d_k \times k \times k$, where $d_k$ is the depth of the kernel and $k$ is the spatial size of the kernel. The size of the cropped video clips for both feeds is 400 $\times$ 400 pixels, which is resized to 120 $\times$ 120 pixels and then randomly cropped to 112 $\times$ 112 pixels. A total of 16 frames in a clip are used as an input of the network for training, which can be denoted as 3 $\times$ 16 $\times$ 112 $\times$ 112. A total of 5 groups of the convolutional layer are used in the network. The size of the convolutional kernel and the extracted feature map from each layer are presented in Figure 2. There are 2 types of residual blocks used in the last 4 convolution layer groups, which are shown at the bottom of Figure 2. The shortcut structure of each block can be expressed as:

$$x_{l+1} = \mathrm{F}(x_l, w_l) + x_l \quad (1)$$

where $x_l$, $w_l$ are the input and weights of the convolutional layer $l$ respectively. $\mathrm{F}(x_l, w_l)$ represents the function where the residual mapping is learned. Batch normalisation (BN) is employed in each convolution layer.

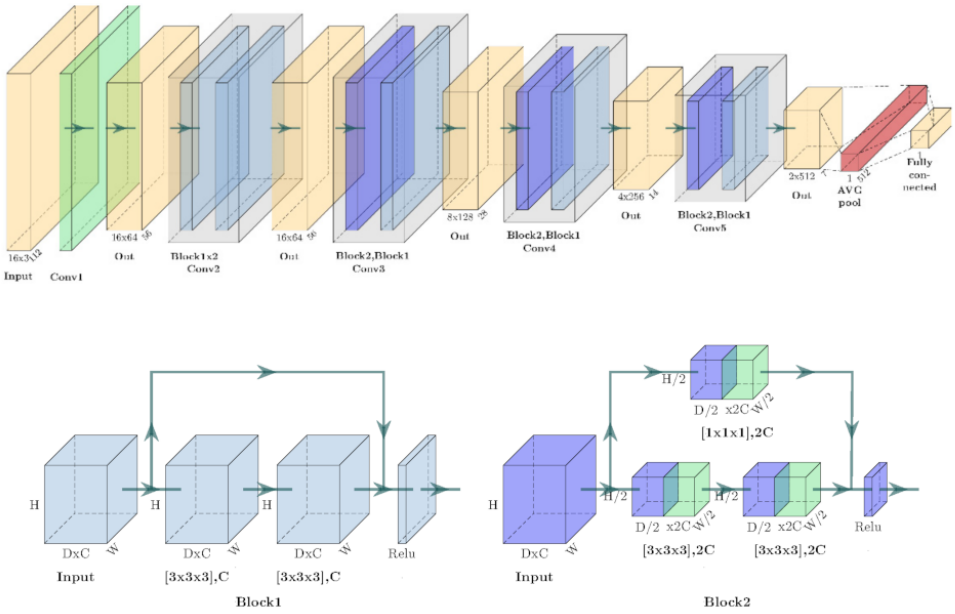


**Figure 2.** The proposed 3D CNN model. There are two types of residual block in this network, which are detailed in the bottom graph and indicated as different colours.

For the training process, Kaiming initialization [46] is applied for weight initialization. The initial learning rate is set as 0.001, which is dynamically reduced when the validation loss stops improving. The loss of a prediction output $x$ in this network can be described as:

$$Loss(x, label) = w_c[label]\left(-x[label] + \log\left(\sum_j e^{(x[j])}\right)\right) \quad (2)$$

where $w_c$ is the weight distribution of the classes in the dataset to improve the data imbalance in a mini-batch; label is the true class of the instance; $j$ is the index of all classes. The losses are averaged for each mini-batch.

The prediction probability of NDA detection based on the driver's head movement is denoted as $P_d$, which has only two states: DA engagement and NDA engagement, denoted as $c_d$ and $c_N$, respectively. The prediction probability for these two classes is therefore presented as $P_d[c_d]$ and $P_d[c_N]$. Two different 3D CNN models have been trained for DA and NDA classification based on hand movement. The prediction probability for these 2 models are denoted as $P_{dc}$ and $P_{Nc}$.

The final prediction scores for all NDA and DA classes, denoted by $S$, can be expressed as:

$$S = S_d \cup S_{\mathrm{N}} \quad (3)$$

where $S_{\mathrm{d}}$ and $S_{\mathrm{N}}$ are the final scores of the DA classification and NDA classification, respectively. The score of a single DA can be expressed as:

$$S_d(i_d) = P_{dc}(i_d)P_d[c_d] \quad (4)$$

where $i_d$ is the index of the DAs.

The score of a single NDA can be expressed as:

$$S_N(i_N) = P_{Nc}(i_N)P_d[c_N] \quad (5)$$

where $i_N$ is the index of the NDAs.

### *2.2. Experiment Design*

Take-over process: Figure 3 plots the design of the take-over process in a trial. During a trial, the vehicle was driving automatically initially while the participant was required to engage in a certain type of NDA or check the road. Then, the take-over process started after a lateral offset was implemented to the vehicle. The lateral error is defined as the distance between the vehicle position and the closest point on the path. After a lateral offset is implemented, the vehicle is in an improper position on the road. Then, an acoustic signal was given to the participant as a take-over request (TOR), which requests he/her to take control of the vehicle and bring it back to the right position. In Figure 3, $T_1$ indicates the time needed for the driver to put her/his hand on the steering wheel. To achieve a safe and smooth take-over transition, a haptic torque was implemented to help the driver guide the vehicle to the reference route. The haptic torque was engaged as soon as the driver applies the torque to the wheel and gradually fades away. After the lateral error achieves the maximum value, the vehicle returns to the reference route. A threshold of the safety distance is defined, which indicates that the control transition is finished, and the driver could achieve safe manual driving afterwards. In this study, the threshold was set as 0.7 m, which is the maximum lateral error to keep the vehicle inside the lane. In Figure 3, $T_2$ refers to the time needed from TOR to the time when the vehicle arrives at the threshold, which is considered as a criterion to evaluate the take-over performance in this study.

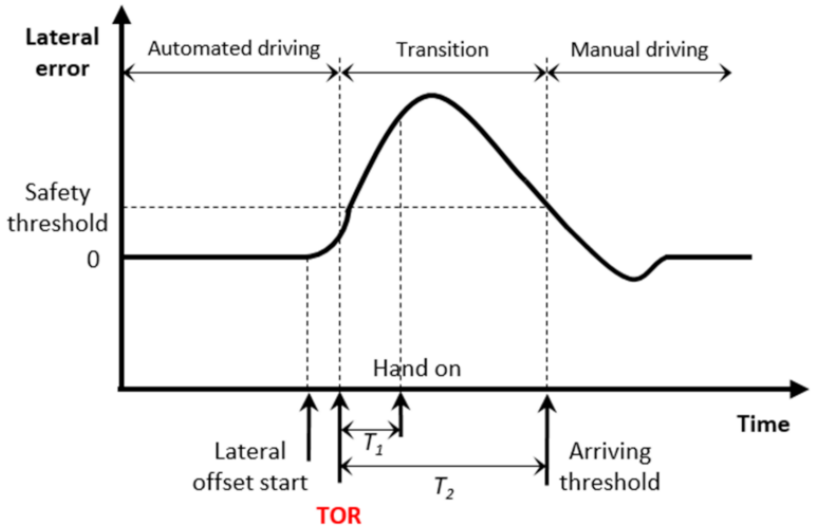


**Figure 3.** Concept of the take-over process.

Track and take-over scenarios: The testing track is a two-lane road with a mini roundabout, as shown in Figure 4. The start point is highlighted with green colour. In the odd loop, the vehicle enters from 1 into the roundabout and leaves from 3. Then, it enters from 4 and leaves from 2. In the even loop, the vehicle enters from 1 and leaves the roundabout from 4. Then, it enters the roundabout from 3 and leaves from 2. The TOR signal was issued at specific points on the track to avoid the area around the roundabout for safety concerns. The lateral offset was set as 1.5 m with a small variation in the real trials. The maximum speed of the vehicle was set as 30 mph. The interval between TORs was randomly selected from the range of 5–9 min.

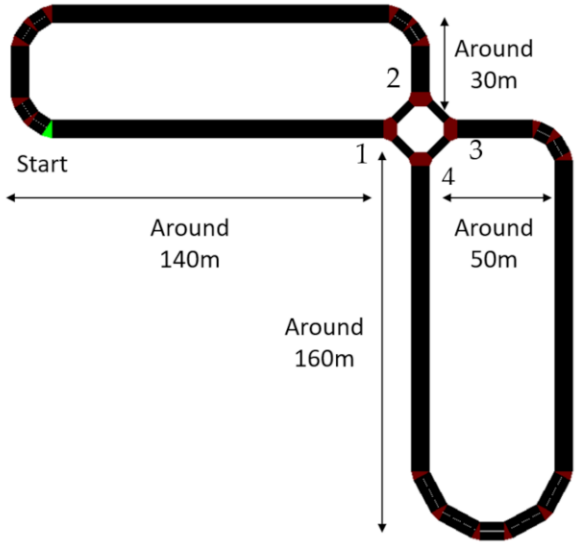


**Figure 4.** Sketch map of the track.

NDAs and DAs: In this study, 4 types of NDA and 2 types of DAs were investigated. A tablet was used for the engagement of NDAs, which include reading news, watching videos, playing games, and answering questionnaires. For reading news, the participants were required to read some articles from BBC News. For watching videos, the participants were asked to watch some videos from YouTube. For playing games, the participants were required to play Temple Run. For the NDA of answering questionnaires, the participants were required to complete a questionnaire, which comprised some objective and subjective questions about this experiment. The DAs considered in this study are road checking and driving. In the experiment, each participant completed 7 trials, including 4 trials for 4 types of NDA respectively and 1 trial of watching the road with no NDA engagement. For the remaining 2 trials, 2 activities were randomly selected from the 5 activities mentioned above. The order of activities for each participant was randomized.

Participants: 14 participants (12 male and 2 female; aged 24–30) participated in this experiment. A valid UK driving license was required. None of them had driving experience with high-level automated driving vehicles.

Video acquisition and dataset pre-processing: 2 Garmin Virb Action Cameras were used to capture the driver's behaviour inside the vehicle cabin in this experiment. Both cameras had a spatial resolution set at 1920 × 1440 pixels and a temporal resolution set at 30 frames per second (fps). As shown in Figure 5, Camera 1 extracts the driver's head movement and facial information and detects if the driver is checking the road or engaging with NDAs. Camera 2 captures the driver's hand movement when interacting with the tablet or the primary steering wheel, which was mounted on the vehicle's roof between two front seats.

In the dataset for the NDA recognition framework, a single instance, denoted by $I$, contains a pair of synchronised frame stacks ($I_f$, $I_r$) from Camera 1 and Camera 2, respectively. The recorded video from each camera was split into several clips. There are 48 frames in each clip, where 16 adjacent frames were randomly picked and used as an input instance $I_f$ or $I_r$. A total of 3624 instances for 6 classes were extracted from the videos of the 14 participants. Specifically, for the 2 DAs, the driving instances were extracted from the videos of the take-over process at the end of an NDA trial. The road-checking instances were extracted from the videos of the watching road trial and the driver's road checking behaviour in an NDAs trial. For every participant, there are around 40 instances for each NDA class and each DA class. The data of 10 participants were randomly selected and used for training, the data of 2 participants were used for validation, and the remaining

2 participants were used for testing. In summary, a total of 2598, 467, and 559 instances were used for training, validation, and testing, respectively.

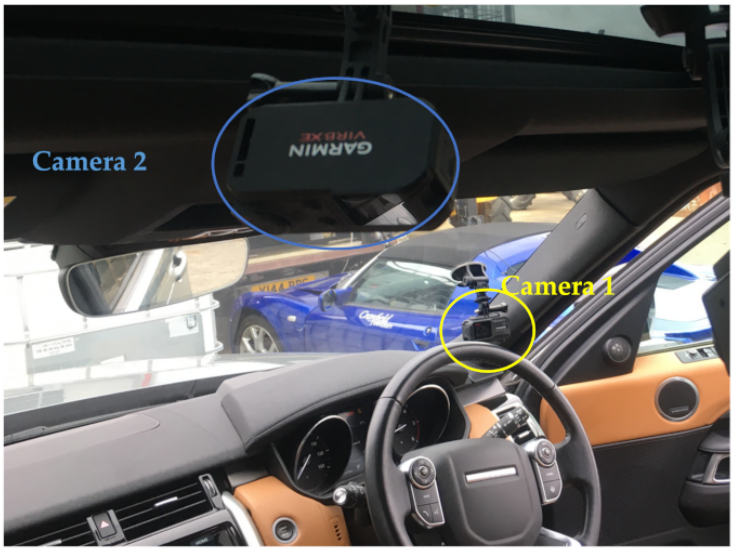


**Figure 5.** A illustration of the two cameras inside the vehicle.

### *2.3. Vehicle Setting*

Vehicle Modification: An instrumented Land Rover Discovery 5 was employed as the testing bed, which was modified to accommodate both autonomous and human driving. An electric motor, operating on the steering column, was used for steering and another electric motor was used to control the throttle pedal position. Braking was modified using a pneumatic actuator on the brake pedal. To ensure safety, a second steering wheel and a set of pedals were added in the back seat, shown in Figure 6, which allows a safety driver to intervene and override the autonomous system. For the path following, the pure pursuit algorithm was used to generate the reference steering angle. The rear steering wheel was controlled using the reference steering angles and the front wheel follows the rear wheel.

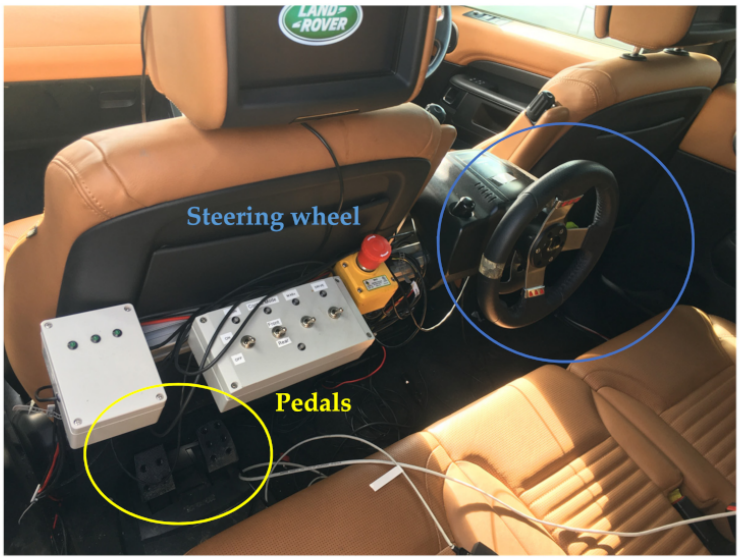


**Figure 6.** A view of the modified vehicle, where the rear steering wheel and pedal have been highlighted.

Vehicle control and data acquisition: The OXTS RT1003 (Oxford Technical Solutions Limited, Oxfordshire, United Kingdom) with RTK GPS system provides the global vehicle position with an accuracy of 2 cm and the heading angle with an accuracy of less than 1 degree. The data of vehicle status were recorded in the Microautobox I (dSPACE, Paderborn, Germany) at a sampling rate of 1 k Hz. The data include driver steering torque, autonomous steering torque, vehicle position and heading, vehicle velocity, steering angle, and take-over signal. The path was recorded beforehand at a sampling rate of 1 k Hz and then resampled by the linear interpolation to a spatial accuracy of 0.2 m.

A certain torque threshold was applied for hand-on-wheel time ($T_1$) detection. It was experimentally determined to avoid false take-over detection due to sensor noise. An instance of the driver's torque during a take-over process is shown in the top plot of Figure 7. The corresponding vehicle route is presented in the bottom plot of Figure 7.

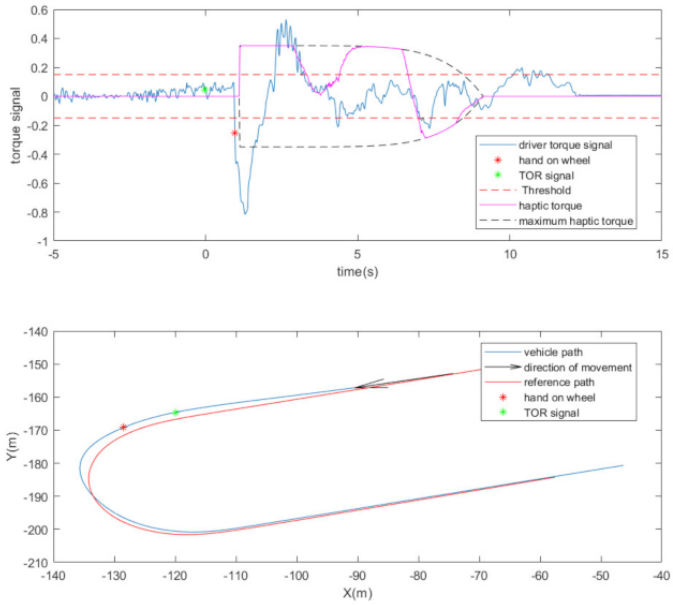


**Figure 7.** Top plot presents the driver's torque and the haptic torque for 1 instance. Bottom plot presents the corresponding vehicle movement in the track.

After the driver takes control of the steering wheel, the vehicle provides haptic cues to the driver, in the form of torque on the steering wheel, to increase the driver's awareness of the environment. The haptic decays over a certain amount of time and eventually reaches 0 to give the driver full control. The value of the torque is calculated using

$$\tau_{haptic}(t) = K_t(t) K_p \left( \delta - \delta_{ref} \right) \tag{6}$$

where $\delta$ is the vehicle steering angle; $\delta_{ref}$ is the reference steering angle calculated by the path following algorithm; $K_p$ is a constant gain, and $K_t(t)$ is a decaying gain which is a function of time starting from 1 and reaching 0 at the end of the take-over period. The decaying profile is represented as the black dot line in the top plot of Figure 7. The decaying duration chosen for this experiment was 8 s. The torque value is normalized between $-1$ and 1, where 1 indicates the maximum torque of the electric motor in one direction and $-1$ indicates the maximum torque in another direction. The maximum amplitude of the torque was a tuning parameter. Each participant tried two of three pre-set values randomly in the trials: 0.35, 0.45, 0.55.

## 3. Results

### *3.1. Activity Classification*

The confusion matrices of the proposed two-feeds 3D CNN models are presented in Figure 8. All six activities are abridged as a term for the convenience of result presentation. Check and Drive refer to two DAs, which are road checking and driving. Game, Ques, Read, and Watch refer to the four kinds of NDA, which are playing games, answering questionnaires, reading news, and watching videos, respectively. For the NDA detection based on the driver's head movement (Figure 8a), the precision and recall of both classes are over 95%. The accuracy of the NDA detection is 97.14%, as shown in Table 1. The DA classification (Figure 8b) also shows a high precision and recall (>90%) and the accuracy is 95.51% (Table 1). A total of 8.4% driving instances were misclassified as road checking, which is due to the hand gesture where sometimes the driver holds the tablet in one hand and controls the steering wheel with another hand during the take-over process. As shown in the confusion matrix for NDA classification (Figure 8c), among all these NDAs, answering questionnaires achieves the best performance (86.5% for precision and 96.0% for recall). The lowest recall value is from playing games, which is 75.5%. The main contribution of the false negative is from watching videos, which is due to the limited hand movement in some game engagement instances, and comparing with the other two NDAs, watching videos has similar spatial information with playing games. The highest value of precision is from reading news (95.7%), while the recall is only 77.9%. Answering questionnaires and watching videos are the main contributions of the false negative. The prediction of reading news could be tricky, because sometimes the limited hand gesture change could lead to confusion with watching videos. Sometimes, the frequent movement required for turning pages could be misclassified as answering questionnaires since the spatial background could be similar between these two classes. From the final fusion matrix, shown in Figure 8d, it can be observed that the driving class achieves high values in both precision and recall. Since the results are obtained by combining the 3 models, the value of each class shows a similar trend with the value in the separate model mentioned above. The total accuracy of the final prediction is 85.87%.

**Table 1.** Performance of 3D CNN model.

| Term | NDA Detection | DA Classification | NDA Classification | Final Prediction |
|---|---|---|---|---|
| Accuracy | 97.14% | 95.51% | 85.56% | 85.87% |
| Weighted F1 score | 97.14% | 95.49% | 85.46% | 85.88% |

### *3.2. Road-Checking Behaviour Analysis*

The road-checking behaviour performed during the NDA engagement has been extracted by the proposed NDA recognition framework and shown in Table 2. The checking period is determined as the division of the total time duration of the NDA engagement trials by the total number of instances of road-checking behaviour in these trials. The motivation of the road-checking behaviour is manually inferred from the recorded videos, which has been divided into four classes, which are bumping, approaching junctions, breakpoint, and other. Bumping refers to the checking behaviour caused by the vehicle vibration that happened on the uneven road surface. Approaching junctions refers to the checking behaviour whereby the driver glanced as the vehicle approached the roundabout and turning. For breakpoint, it indicates the checking behaviour, which is performed in a short break during the NDA engagement. For example, the driver could check the surrounding environment after she/he finished watching a short clip or a round of the game. Other includes the motivations which are different to the above-mentioned class, such as illumination changing or regular checking, etc.

**Figure 8.** Confusion matrices of the 3 developed 3D CNN models and the fused results. The precision and recall for each class are presented in the bottom and right of the figure, respectively.

**Table 2.** Road-checking behaviour evaluation.

| NDAs | Checking Period (s) | Percentage of Checking for Corresponding Motivation | | | |
|---|---|---|---|---|---|
| | | Bumping | Approaching Junctions | Breakpoint | Others |
| Watching videos | 37.10 | 19.88% | 52.05% | 5.85% | 22.22% |
| Reading news | 51.64 | 16.78% | 51.75% | 7.69% | 23.78% |
| Playing games | 79.13 | 3.61% | 26.50% | 59.04% | 10.84% |
| Answering questionnaires | 123.00 | 18.18% | 50.00% | 13.64% | 18.18% |

From Table 2, it can be seen that the checking period of watching videos is the shortest (37.10 s). The period of reading news is slightly higher at 51.64 s. As the passive NDAs, both NDAs show a similar proportion of motivation. Approaching junction is the main motivation, which is above 50%. Playing games shows a relatively longer checking period (79.13 s) than the passive NDAs, where breakpoint (59.04%) is the dominant motivation. Answering questionnaires has the least road-checking behaviour, mostly once or twice during a single trial. The motivation shows a similar trend with the passive NDAs, but with a higher proportion of the breakpoint (13.64%). For the passive NDAs, the driver performs more frequent road-checking behaviour compared to the active NDAs. This suggests that the driver pays less attention to the engagement under this type of NDAs. The main motivations are approaching junctions and bumping. This suggests that the driver is more sensitive to the change of the vehicle state, such as velocity change, vibration, or turning.

Such checking behaviour is important for safety control transition if the TOR is given when the vehicle state is changing. The results of the active NDAs engagement show that the driver performs much less road-checking behaviour, particularly for the engagement of answering questionnaires. For playing games, the checking behaviour mainly happens in breakpoint, which suggests the driver is at a high attention level and not sensitive to the environment change during the engagement. Therefore, for the passive NDA, achieving a high-quality control transition could be more challenging for the driver due to the limited situational awareness.

### *3.3. Take-Over Performance*

The take-over performance is presented and evaluated in this section. Figure 9 shows the driver's hand-on-wheel time ($T_1$) against five activities. The shortest mean value of $T_1$ is around 1.3 s, which is NoTask. For the remaining four NDAs, the average value is in the range of 1.9–2.6 s. Answering questionnaires shows a relatively shorter $T_1$ than other NDAs. The maximum lateral error is presented in Figure 10. It can be observed that the value for each NDA and NoTask are similar, the average value is around 2.8 m. In the experiment, the TOR signal was given after the lateral error achieved 1.5 m, while most of the drivers could control the vehicle within a maximum 3.5 m lateral error. However, the impact of the NDAs engagement on the take-over performance mainly presented after the maximum lateral error was achieved. From Figure 11, the baseline of time needed to achieve the safe position ($T_2$) is around 4.16 s for NoTask. The $T_2$ for NDAs engagement is at least 0.5 s more, which suggests the engagement of NDAs could increase the time that the vehicle stays in the dangerous position on the road. The mean and standard deviation of $T_2$ for each activity are presented in Table 3. For the passive NDAs, the mean values of watching videos and reading news, as the passive NDAs, are 4.74 s and 4.96 s respectively, which are higher than the active NDAs, answering questionnaires and playing games (5.45 s and 5.43 s respectively). The standard deviation of the NDAs is higher than the NoTask, which suggests higher individual differences in terms of take-over performance in NDAs engagement. From Figures 9 and 11, it can be observed that the driver under active NDAs engagement requests more time to drive the vehicle back to the safe position, which suggests a higher mental demand or workload during such NDAs engagement. After receiving the TOR signal, the driver needs more time to build awareness of the surrounding environment and it is more challenging for them to switch to the take-over process from the NDA engagement.

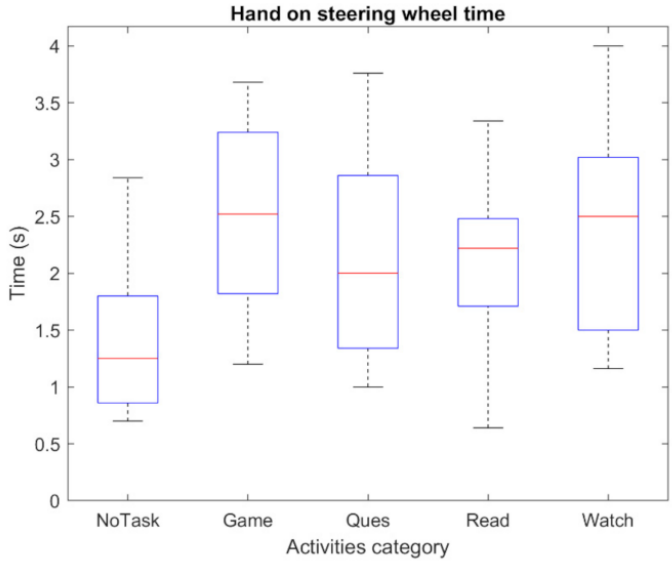


**Figure 9.** The hand-on-wheel time performance. NoTask refers to the performance in watching road trial.

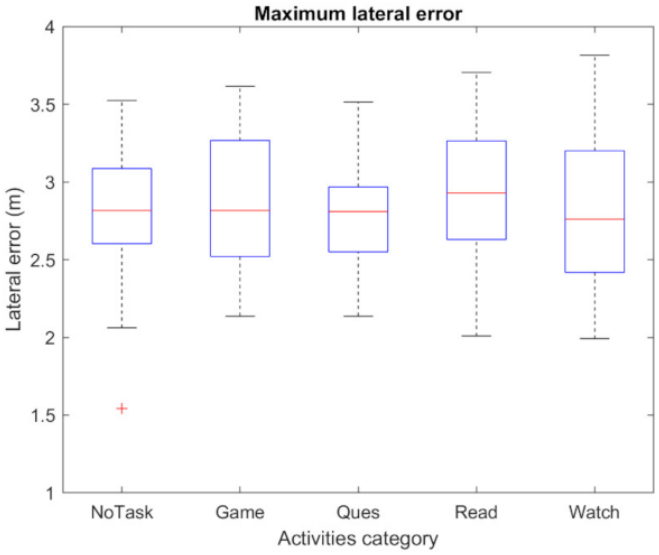


**Figure 10.** Maximum lateral error achieving. An outlier is presented with the plus sign which represents a high concentration level of the driver during the road checking task.

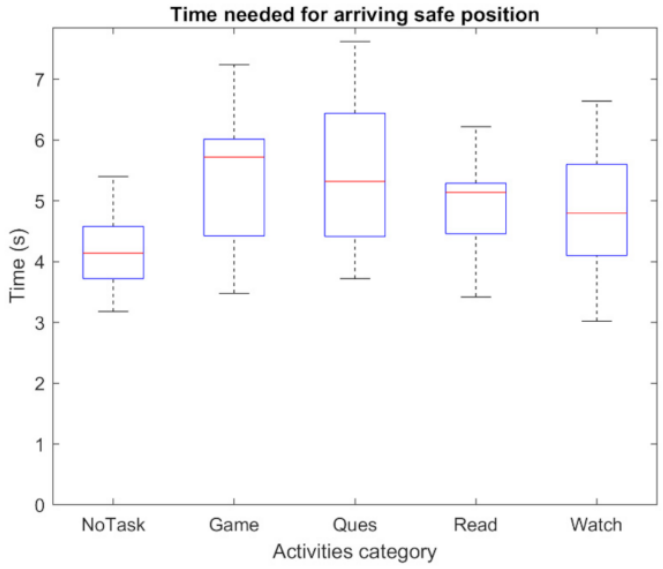


**Figure 11.** Time cost for the vehicle back to the safe position.

**Table 3.** Time to threshold for all activities.

| Time to Threshold | Activities | | | | |
|---|---|---|---|---|---|
| | **No Task** | **Watch** | **Read** | **Ques** | **Game** |
| Mean (s) | 4.16 | 4.74 | 4.96 | 5.45 | 5.43 |
| Standard deviation (s) | 0.67 | 1.12 | 0.87 | 1.23 | 1.14 |

Table 4 presents the $T_2$ under different levels of haptic feedback. For a low level of haptic torque, the mean value of $T_2$ is 5.32 s, which is the lowest among all the evaluated levels. The standard deviation is 1.12 s, which suggests that all the participants have higher tolerance at this level of haptic torque assistance. It can be seen that the increase of the torque level could result in the decrease of the mean value of $T_2$, which means a higher level of haptic torque could support the driver to reduce $T_2$ and improve their take-over performance. However, the standard deviation increases (1.55 s for medium level and 1.32 s

for high level). This suggests that some of the participants could distrust and resist the higher level of haptic torque and take a longer $T_2$.

**Table 4.** Time to threshold for different haptic torque levels.

| Time to Threshold | Haptic Torque Level | | |
|---|---|---|---|
| | **Low** | **Medium** | **High** |
| Mean (s) | 5.32 | 4.97 | 4.83 |
| Standard deviation (s) | 1.12 | 1.55 | 1.32 |

## 4. Conclusions

Achieving a safe control transition is one of the most important challenges in level 3 automation systems and is influenced by many factors, where the driver's mental state and driving-environment awareness before take-over play an important role. We proposed a two-feed 3D CNN-based NDA recognition system which can automatically detect and classify the driver's NDAs engagement and DA activities with high accuracy. It has been demonstrated that both head and hand movement are crucial for achieving this target. This study further investigated the implication of the NDAs engagement on both perspectives of roading checking behaviour and take-over performance. Based on the investigation, a category method has been proposed to group the NDAs, which aims to extend the application of this study on a wide range of NDAs. Moreover, the effectiveness of the steering wheel haptic assistance system for the take-over process has been evaluated.

For the investigation of road checking behaviour, the driver always performs such behaviour during the engagement of NDAs to ensure driving safety. There is less road checking behaviour under the active NDAs engagement. The motivation study shows the driver mainly checks the road in the breakpoint and is less sensitive to the change of the vehicle state, which suggests that the driver paid more attention to the activity and has less awareness of the driving environment. Since the lack of observation could be dangerous, the driver should be reminded to monitor their surroundings to improve situation awareness when they engage in this kind of NDA for a long period without road-checking behaviour. From the take-over performance point of view, the engagement of NDAs leads to a negative effect (longer $T_2$). The engagement of active NDAs could demand even more time. Furthermore, haptic torque assistance could improve the take-over performance, as evidenced by decreasing $T_2$. However, a higher level of haptic torque could result in the driver's resistance.

In summary, this investigation helps us develop a deeper understanding of the implication of the driver's behaviour on the control transition in conditional automation, which further helps the design of HMI and take-over strategies to accomplish a safe take-over. The type of NDA determines the level of the driver's mental demand, which further affects their situation awareness and take-over performance. The observed results also suggest that the take-over process could benefit from highly frequent road-checking and haptic feedback assistance. The existing HMI design only considers the type of NDAs in the take-over process from the perspective of the take-over request modality. From this study, the road-checking behaviour of the driver during NDAs engagement also matters for reducing the take-over time. An alert system for checking the surrounding environment should be considered for further HMI design.

**Author Contributions:** Conceptualization, L.Y., Y.Z. and Y.X.; methodology, L.Y. and M.B.S.; software, L.Y.; validation, Y.X., C.L. and J.B.; formal analysis, L.Y.; investigation, L.Y.; resources, M.B.S.; data curation, M.B.S.; writing—original draft preparation, L.Y. and M.B.S.; writing—review and editing, Y.X., C.L. and Y.Z.; visualization, L.Y.; supervision, J.B., C.L. and Y.Z.; project administration, Y.Z.; funding acquisition, J.B. and Y.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by Cranfield's EPSRC Impact Acceleration Account EP/R511511/1.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References


1. SAE International Standard J3016_201806*Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*, SAE: Warrendale, PA, USA, 2018.
2. National Transportation Safety Board. *Collision Between a Car Operating with Automated Vehicle Control Systems and a Tractor-Semitrailer Truck*; National Transportation Safety Board: Washington, DC, USA, 2017.
3. National Transportation Safety Board. *Collision Between a Sport Utility Vehicle Operating with Partial Driving Automation and a Crash Attenuator, Mountain View, California, March 23, 2018*; National Transportation Safety Board: Washington, DC, USA, 2018.
4. de Winter, J.C.F.; Happee, R.; Martens, M.H.; Stanton, N.A. Effects of adaptive cruise control and highly automated driving on workload and situation awareness: A review of the empirical evidence. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *27*, 196–217. [CrossRef]
5. Dogan, E.; Rahal, M.-C.; Deborne, R.; Delhomme, P.; Kemeny, A.; Perrin, J. Transition of control in a partially automated vehicle: Effects of anticipation and non-driving-related task involvement. *Transp. Res. Part F Traffic Psychol. Behav.* **2017**, *46*, 205–215. [CrossRef]
6. Yoon, S.H.; Ji, Y.G. Non-driving-related tasks, workload, and takeover performance in highly automated driving contexts. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *60*, 620–631. [CrossRef]
7. Aksjonov, A.; Nedoma, P.; Vodovozov, V.; Petlenkov, E.; Herrmann, M. Detection and Evaluation of Driver Distraction Using Machine Learning and Fuzzy Logic. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 2048–2059. [CrossRef]
8. Osman, O.A.; Hajij, M.; Karbalaieali, S.; Ishak, S. A hierarchical machine learning classification approach for secondary task identification from observed driving behavior data. *Accid. Anal. Prev.* **2019**, *123*, 274–281. [CrossRef]
9. Martin, M.; Popp, J.; Anneken, M.; Voit, M.; Stiefelhagen, R. Body Pose and Context Information for Driver Secondary Task Detection. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2015–2021.
10. Xing, Y.; Lv, C.; Zhang, Z.; Wang, H.; Na, X.; Cao, D.; Velenis, E.; Wang, F.-Y. Identification and analysis of driver postures for in-vehicle driving activities and secondary tasks recognition. *IEEE Trans. Comput. Soc. Syst.* **2018**, *5*, 95–108. [CrossRef]
11. Li, N.; Busso, C. Detecting drivers' mirror-checking actions and its application to maneuver and secondary task recognition. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 980–992. [CrossRef]
12. Jin, L.; Guo, B.; Jiang, Y.; Wang, F.; Xie, X.; Gao, M. Study on the impact degrees of several driving behaviors when driving while performing secondary tasks. *IEEE Access* **2018**, *6*, 65772–65782. [CrossRef]
13. Yang, L.; Dong, K.; Dmitruk, A.J.; Brighton, J.; Zhao, Y. A dual-cameras-based driver gaze mapping system with an application on non-driving activities monitoring. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 4318–4327. [CrossRef]
14. Yoon, S.H.; Kim, Y.W.; Ji, Y.G. The effects of takeover request modalities on highly automated car control transitions. *Accid. Anal. Prev.* **2019**, *123*, 150–158. [CrossRef]
15. Clark, H.; Feng, J. Age differences in the takeover of vehicle control and engagement in non-driving-related activities in simulated driving with conditional automation. *Accid. Anal. Prev.* **2017**, *106*, 468–479. [CrossRef] [PubMed]
16. Li, S.; Blythe, P.; Guo, W.; Namdeo, A. Investigating the effects of age and disengagement in driving on driver's takeover control performance in highly automated vehicles. *Transp. Plan. Technol.* **2019**, *42*, 470–497. [CrossRef]
17. 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. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2014**, *58*, 2063–2067. [CrossRef]
18. Scharfe, M.S.L.; Zeeb, K.; Russwinkel, N. The impact of situational complexity and familiarity on takeover quality in uncritical highly automated driving scenarios. *Information* **2020**, *11*, 115. [CrossRef]
19. Wu, C.; Wu, H.; Lyu, N.; Zheng, M. Take-over performance and safety analysis under different scenarios and secondary tasks in conditionally automated driving. *IEEE Access* **2019**, *7*, 136924–136933. [CrossRef]
20. Wandtner, B.; Schömig, N.; Schmidt, G. Effects of non-driving related task modalities on takeover performance in highly automated driving. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2018**, *60*, 870–881. [CrossRef] [PubMed]
21. Jeong, H.; Liu, Y. Effects of non-driving-related-task modality and road geometry on eye movements, lane-keeping performance, and workload while driving. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *60*, 157–171. [CrossRef]
22. Kim, J.; Kim, H.-S.; Kim, W.; Yoon, D. Take-over performance analysis depending on the drivers' non-driving secondary tasks in automated vehicles. In Proceedings of the 2018 International Conference on Information and Communication Technology Convergence (ICTC), Jeju Island, Korea, 17–19 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1364–1366.
23. Zeeb, K.; Buchner, A.; Schrauf, M. Is take-over time all that matters? *The impact of visual-cognitive load on driver take-over quality after conditionally automated driving. Accid. Anal. Prev.* **2016**, *92*, 230–239. [CrossRef]
24. Petermeijer, S.; Doubek, F.; de Winter, J. Driver response times to auditory, visual, and tactile take-over requests: A simulator study with 101 participants. In Proceedings of the 2017 IEEE International Conference on Systems, Man and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1505–1510.

25. Kim, H.; Kim, W.; Kim, J.; Yoon, D. A study on the control authority transition characteristics by driver information. In Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1562–1563.
26. Choi, D.; Sato, T.; Ando, T.; Abe, T.; Akamatsu, M.; Kitazaki, S. Effects of cognitive and visual loads on driving performance after take-over request (TOR) in automated driving. *Appl. Ergon.* **2020**, *85*, 103074. [CrossRef]
27. Bueno, M.; Dogan, E.; Hadj Selem, F.; Monacelli, E.; Boverie, S.; Guillaume, A. How different mental workload levels affect the take-over control after automated driving. In Proceedings of the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC), Rio de Janeiro, Brazil, 1–4 November 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2040–2045.
28. Petersen, L.; Robert, L.; Yang, X.J.; Tilbury, D. Situational awareness, driver's trust in automated driving systems and secondary task performance. *SAE Int. J. Connect. Autom. Veh.* **2019**, *2*, 129–141. [CrossRef]
29. Wandtner, B.; Schmidt, G.; Schoemig, N.; Kunde, W. Non-driving related tasks in highly automated driving—Effects of task modalities and cognitive workload on take-over performance. In Proceedings of the AmE 2018—Automotive meets Electronics, 9th GMM-Symposium, Dortmund, Germany, 7–8 March 2018; pp. 1–6.
30. Sivak, M.; Schoettle, B. *Motion Sickness in Self-Driving Vehicles*; University of Michigan, Transportation Research Institute: Ann Arbor, MI, USA, 2015.
31. Petermeijer, S.M.; Abbink, D.A.; Mulder, M.; de Winter, J.C.F. The effect of haptic support systems on driver performance: A literature survey. *IEEE Trans. Haptics* **2015**, *8*, 467–479. [CrossRef]
32. Wan, J.; Wu, C. The effects of vibration patterns of take-over request and non-driving tasks on taking-over control of automated vehicles. *Int. J. Hum. Comput. Interact.* **2018**, *34*, 987–998. [CrossRef]
33. Lv, C.; Wang, H.; Cao, D.; Zhao, Y.; Auger, D.J.; Sullman, M.; Matthias, R.; Skrypchuk, L.; Mouzakitis, A. Characterization of driver neuromuscular dynamics for human–automation collaboration design of automated vehicles. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 2558–2567. [CrossRef]
34. Petermeijer, S.M.; de Winter, J.C.F.; Bengler, K.J. Vibrotactile Displays: A survey with a view on highly automated driving. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 897–907. [CrossRef]
35. Lv, C.; Wang, H.; Cao, D.; Zhao, Y.; Sullman, M.; Auger, D.J.; Brighton, J.; Matthias, R.; Skrypchuk, L.; Mouzakitis, A. A novel control framework of haptic take-over system for automated vehicles. In Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, 26–30 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1596–1601.
36. Lv, C.; Li, Y.; Xing, Y.; Huang, C.; Cao, D.; Zhao, Y.; Liu, Y. Human-Machine Collaboration for Automated Vehicles via an Intelligent Two-Phase Haptic Interface. *arXiv* **2020**, arXiv:2002.03597.
37. Vicente, F.; Huang, Z.; Xiong, X.; De la Torre, F.; Zhang, W.; Levi, D. Driver gaze tracking and eyes off the road detection system. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2014–2027. [CrossRef]
38. Fridman, L.; Langhans, P.; Lee, J.; Reimer, B. Driver gaze region estimation without use of eye movement. *IEEE Intell. Syst.* **2016**, *31*, 49–56. [CrossRef]
39. Naqvi, R.A.; Arsalan, M.; Batchuluun, G.; Yoon, H.S.; Park, K.R. Deep learning-based gaze detection system for automobile drivers using a NIR camera sensor. *Sensors* **2018**, *18*, 456. [CrossRef] [PubMed]
40. Yang, L.; Dong, K.; Ding, Y.; Brighton, J.; Zhan, Z.; Zhao, Y. Recognition of visual-related non-driving activities using a dual-camera monitoring system. *Pattern Recognit.* **2021**, *116*, 107955. [CrossRef]
41. Yang, L.; Yang, T.; Liu, H.; Shan, X.; Brighton, J.; Skrypchuk, L.; Mouzakitis, A.; Zhao, Y. A refined non-driving activity classification using a two-stream convolutional neural network. *IEEE Sens. J.* **2020**, *21*, 15574–15583. [CrossRef]
42. Xing, Y.; Lv, C.; Wang, H.; Cao, D.; Velenis, E.; Wang, F.-Y. Driver activity recognition for intelligent vehicles: A deep learning approach. *IEEE Trans. Veh. Technol.* **2019**, *68*, 5379–5390. [CrossRef]
43. Ji, S.; Xu, W.; Yang, M.; Yu, K. 3D Convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 221–231. [CrossRef] [PubMed]
44. Hara, K.; Kataoka, H.; Satoh, Y. Learning spatio-temporal features with 3D residual networks for action recognition. In Proceedings of the 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, 22–29 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 3154–3160.
45. Du, X.; Li, Y.; Cui, Y.; Qian, R.; Li, J.; Bello, I. Revisiting 3D ResNets for video recognition. *arXiv* **2021**, arXiv:2109.01696.
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1026–1034.

https://dspace.org/

# The identification of non-driving activities with associated implication on the take-over process

**Yang, Lichao; Semiromi, Mahdi Babayi; Xing, Yang; Lv, Chen; Brighton, James ...**

**2021-12-22**
MDPI
https://doi.org/10.3390/s22010042
Yang L, Semiromi MB, Xing Y, et al., (2021) The identification of non-driving activities with associated implication on the take-over process, Sensors, Volume 22, Issue 1, December 2021, Article number 42

Downloaded from a DSpace repository. https://dspace.org/