MASTER

Reflection of Driver Drowsiness in Physiological and Vehicular Signals

Sourras, Konstantinos

Award date:
2023

Eindhoven, 06/04/2023

Reflection of Driver Drowsiness in Physiological and Vehicular Signals

by Konstantinos Sourras

identity number 1009093

in partial fulfilment of the requirements for the degree of

Master of Science
in Human-Technology Interaction

Supervisors:
Joyce Westerink
Karin Smolders

Abstract

Drowsy driving is one of the major contributors to traffic accidents and fatalities. Multiple studies have been performed and tried to determine factors that can allow for the early detection of drowsiness and, ultimately, the prevention of accidents. The aim of this study is to look into which physiological measures, such as Heart Rate, Respiration Rate and Heart Rate Variability (RMSSD) can potentially be used in combination with the currently used approaches to Driver Drowsiness Detection systems, which mainly involve measures obtained from the vehicle itself, to further enhance the detection algorithms. For the purpose of this research, an already existing dataset is used, consisting of physiological, vehicular and subjective measures of 19 male participants in a driving simulator. In order to assess if the introduction of physiological measures to the vehicular based models can have additional value, step-wise mixed models are created and the explained variance of the vehicular signal only model is used as a comparison metric. The main findings suggest that both Respiration Rate and Heart Rate can be viable predictors for driver drowsiness, while Heart Rate Variability (RMSSD) does not seem to have a significant effect, which is in line with previous findings. Lastly, some suggestions for future research in the development of Driver Drowsiness Detection systems are proposed and discussed.

Reflection of Driver Drowsiness in Physiological and Vehicular Signals

Introduction

Drowsiness is defined as the state lying between wakefulness and sleep and is related the the inclination to sleep (Slater, 2008). Operating a vehicle while being sleepy or drowsy is commonly referred to as drowsy driving. There are a multitude of factors that can lead to drowsiness, such as sleep deprivation or disorders, medication and drug effects, and lifestyle factors (Lagarde & Batejat, 1994; Sunwoo et al., 2017). Drowsiness affects the driver's ability to concentrate on the road (Ngxande, Tapamo, & Burke, 2017), decreases their response time and affects their ability to process information and make good decisions (Sunwoo et al., 2017). Drowsy and fatigued driving is believed to be one of the main contributors to traffic accidents around the world and was estimated to be a contributing factor to 15-20% of crashes in the Netherlands (SWOV, 2019). This implies the usefulness of a system that can detect early signs of drowsiness during driving and alert the driver in time, in order to stimulate their alertness or awareness of their mental state.

In the literature, various solutions are reported to determine the drowsiness of drivers. The proposed solutions, so far, involve five different kinds of information on which the measurements are based. The types of information used in the proposed solutions are listed below and further discussed in the subsequent sections:

- behavior-based,
- physiology-based,
- vehicular-based,
- subjective measures and
- hybrid-based measure approaches (Bajaj, Kumar, & Kaushal, 2021).

In the following subsections, a summary of the related work for each approach is presented.

Behavior-based Approaches

A driver does not directly transit from being alert to drowsy and there are behavioral signs that accompany the transition. Nordbakke and Sagberg (2007) reported that such signs are observable and can be captured by cameras. The behaviors include changes in eye blinking patterns, yawning frequency, changes in posture and facial expressions, and they become more evident with increasing drowsiness. There are three main categories of behavior-based techniques depending on which bodily areas are being observed, namely: mouth, head and eyes (Albadawi, Takruri, & Awad, 2022).

In their review Hussein, Salman, Miry, and Subhi (2021) report the most used features for each of the three, aforementioned, area-based categories. While considering the eyes of the driver, the blinking frequency, closure duration, PERcentage of eyelid CLOSure (PERCLOS) and Eye Aspect Ration (EAR) are commonly. The features associated with the observation of the mouth and head are yawning frequency and posture, respectively.

Behavior based approaches are reported to have high accuracy rate and are non-intrusive. However, such applications face some limitations: the results are dependent on the quality of the images or videos used, and are affected by external factors (e.g., lighting conditions). Lastly, Gonçalves and Bengler (2015) indicated in their work that one should not exclusively depend on behavioral data to determine the driver's state, because people that perceive themselves as drowsy, might also perform these behaviors, and people that do not show these behaviors, can still be drowsy.

Physiology-based Approaches

Physiology-based signals, also referred to as biological-based in the literature, have been reported to be able to detect drowsiness in an early stage (Bajaj et al., 2021; Chen, Zhang, & Lou, 2019). A multitude of different physiological signals have been used in the literature to define the level of drowsiness for drivers such as electroencephalogram (EEG), electro-oculogram (EoG), electromyogram (EMG), electrocardiogram (ECG) and respiratory signals. One of the added benefits of involving physiological signals in a model, comes from the idea that they are impartial, however, they require special equipment and can be of intrusive nature (i.e., requires electrodes to be placed on the driver).

To the author's knowledge, the most commonly used features explored in the literature involve EEG-based and ECG-based features (i.e., Heart Rate (HR) and Heart Rate Variability (HRV)), while Respiratory signals seem to be studied in a lesser degree when Drowsy Driving Detection (DDD) is concerned. Heart Rate and Respiration Rate indicate the number of heartbeats and full breaths per minute, respectively. Heart Rate Variability, on the other hand, is a measure of variation between consecutive heartbeats in the time domain (Shaffer & Ginsberg, 2017).

Many researchers have explored the relationship between EEG signals, driving and drowsiness, by means of the different frequencies of brain waves being present in different states and have indicated that EEG can predict drowsiness with very high accuracy (Artaud et al., 1994; Finelli, Baumann, Borbély, & Achermann, 2000). As far as EMG is considered, research performed in a driving simulation, suggests that muscles also show signs of fatigue; the muscle movements of the drivers change by means of frequency and amplitude (Balasubramanian & Adalarasu, 2007; Katsis, Ntouvas, Bafas, & Fotiadis, 2004). In later research, muscular fatigue was correlated with drowsiness detection Satti, Kim, Yi, Cho, and Cho (2021), indicating that muscle activity of lower frequency and amplitude were associated with increased drowsiness levels.

Multiple authors report that specific patterns are observable in physiological markers when a driver transits to a drowsy state. More specifically, there is an increase in the Heart Rate of a person and a decrease in their Respiration Rate (Kiashari, Nahvi, Homayounfard, & Bakhoda, 2018; Warwick, Symons, Chen, & Xiong, 2015). The patterns were both observed for females and males, although the duration and magnitude of the decreases and increases were more noticeable in males compared to females (Warwick et al., 2015). In more recent research conducted by (Jo, Kim, & Kim, 2019), where they also investigated how Heart Rate changes while drowsy driving, they found that during normal driving Heart Rate does increase, compared to the average daytime Heart Rate, but decreases when drowsy driving.

Heart Rate Variability (HRV), according to the literature, should be a good indicator of drowsiness (Khushaba, Kodagoda, Lal, & Dissanayake, 2011). However, in their work (Warwick et al., 2015) report that this was not consistent in their series of experiments, which is in line with the findings of van den Berg, Neely, Wiklund, and Landström (2005), namely that HRV does not show a significant change between awake and asleep states and, thus, does not seem like a viable indicator of drowsiness.

Vehicular-based Approaches

State-of-the-art approaches, used in the automotive industry, rely on vehicular based solutions (Doudou, Bouabdallah, & Berge-Cherfaoui, 2020). Vehicular-based approaches aim to detect changes in driving patterns. Commonly used measures are the Steering Wheel Angle or Movement (SWA/M, respectively) and Lane Departure. SWA/M can be tracked by means of a sensor placed on the steering wheel and Lane Departure metrics are measured with an external camera, based on which the position of the car with respect to the street lanes is calculated.

The most common feature derived from the Lane Departure metrics is the Standard Deviation of the Lateral Position (SDLP) (Bajaj et al., 2021) and has been found to increase for higher drowsiness (Ingre, Åkerstedt, Peters, Anund, & Kecklund, 2006). Research involving SWA/M has indicated that the drowsiness level of the driver can be determined by the amount of the small movements (i.e., movements aiming to correct the position of the car, rather than change a lane) based on the observation that the occurrence of micro-corrections becomes reduced while drowsy driving, compared to normal driving (Feng, Zhang, & Cheng, 2009). Additional measures are also utilized by many scholars, examples of such measures are speed, acceleration and pressure on the acceleration pedal (Doudou et al., 2020).

In their review, Sahayadhas, Sundaraj, and Murugappan (2012) report that vehicular-based solutions lead to poor predictors for DDD and, depending on the driving patterns of the driver, might detect the drowsiness too late to prevent an accident. Furthermore, in their work Gwak, Hirao, and Shino (2020) and Li, Li, Li, Cheng, and Shi (2017) indicate that vehicular-based approaches suffer from high false positive rate.

Even though vehicular-based approaches raise some additional challenges, such as the dependence on factors external to the driver (i.e., road infrastructure, weather and illumination conditions), the signals utilized by this kind of approaches can be retrieved from most modern vehicles without the need of additional sensors and can be processed in real time, which is not the case for most of the other approaches. Additionally, although vehicular-based solutions aim to detect changes in driving behavior, they cannot differentiate between drowsiness and other factors (e.g., alcohol or drug consumption) (Bajaj et al., 2021).

Subjective-based Approaches

Subjective based solutions mainly rely on questionnaires that the drivers fill in on their own and are meant to reflect the drowsiness as perceived by the driver themselves. Based on the literature accessed, there is currently no standardized questionnaire being used (Bajaj et al., 2021). However, the Karolinska Sleepiness Scale (KSS) is frequently used (Kaida et al., 2006; Philip et al., 2005), during which the driver/participant indicates their perceived level of sleepiness in a 9-point scale, with 1 indicating that the participant is "*extremely alert*", 5 being neutral; "*neither alert nor sleepy*" and 9 indicating "*Very sleepy, great effort to keep alert, fighting sleep*". Another available questionnaire is the Stanford Sleepiness Scale (SSS), which is a 7-point scale which the driver/participant fill in themselves as well (Hoddes, Zarcone, Smythe, Phillips, & Dement, 1973). Due to both the KSS and SSS questionnaires being filled in at regular intervals, it cannot address sudden changes in sleepiness. Moreover, the fact that the driver has to fill it in; can stimulate the attention of the driver, which does interfere with the measured variable. Note that interfering with the sleepiness of the drivers in a real world scenario is not a problem, as Driver Drowsiness Detection systems aim to do so. However, in experiments where you want to assess the sleepiness of the participants or study the relation between sleepiness and other measurements to get the ground truth, this can be problematic. As a solution to the experimental problem of subjective reporting and the stimulation due to the repetitive nature of these measurements, different approaches can be found in the literature. For instance, scholars have suggested the use of objective criteria/behaviors and trained observers to take over the procedure of rating drowsiness in experimental setups (Wierwille & Ellsworth, 1994).

Hybrid-based Approaches

Hybrid driver drowsiness detection systems combine at least two of behavior-, physiology- and vehicular-based measures aiming to create more reliable and accurate DDD systems. Various attempts have been made which combine different signals from different categories. Based on the overview of Albadawi et al. (2022) the highest accuracy was achieved by the combination of behavior- and vehicular-based measures, followed by a physiological and vehicular approach utilizing signals such as Heart Rate, stress level, Respiration Rate and SW Acceleration. However, as mentioned previously, comparing approaches based on their reported accuracy might not be meaningful, as a different dataset is used for each test and different simulation setups are used. This being said, in a practical situation, there is no proof that one system will outperform another. Even by assuming that both can predict drowsiness to the same degree, there are still considerations that need to be taken into account and a preference over one of the two might be indicated. For instance, as mentioned previously, some vehicular measures make use of cameras to measure the SDLP and are dependent on road infrastructure (i.e., appropriate lane markings) and illumination conditions. The use of cameras on the exterior of cars can, additionally, raise concerns with regards to privacy and, given that there is no option to opt out of it, may be problematic for some. On the contrary, physiological measures (besides some approaches utilizing cameras) can be more impartial to such factors and end up being more accurate during driving sessions where lighting conditions change, or when driving in roads without sufficient road surface markings. This may indicate that the utilization of physiological based measures would be preferred.

Research Question

In the literature, there is are several references discussing various hybrid approaches focusing on the combination of vehicular and physiological measures. However. even in the approaches combining all types of measures (i.e., including all three categories; behavior-, physiology- and vehicular-based), mainly Heart Rate (HR) is considered from the physiology-based features, and respiratory features are not given much attention, although both are reported to be able to detect drowsiness. The absence of such an approach may indicate a potential gap in the literature or might be a result of publication bias. As a result, this research aims to investigate to what extent physiological measures, including Respiration Rate, can be used in combination with the vehicular-based features to describe the driver's drowsiness level. This leads to the following research question:

RQ: *"To what extent can respiratory signals/patterns reflect the drowsiness level of drivers, and to what extent can physiological signals, by means of Heart Rate, Heart Rate Variability and Respiration Rate help to improve drowsiness predictions based on only vehicular measures?"*

Based on the accessed literature, it is expected that taking into consideration physiological information will increase the performance of the models only considering vehicular-based measures. Current approaches only consider driving behavior changes and are susceptible to external factors (i.e., road and light conditions). Adding a physiological measurement could lead to a more "universal" solution.

Furthermore, the following hypotheses will be tested:

H1: *"Respiration Rate can be used to predict the drowsiness level of the driver."*
H2a: *"Combining the Respiration Rate (RR) and vehicular-based signals will represent the drowsiness level of the driver more accurately, compared to only vehicular-signals."*
H2b: *"Combining the Heart Rate (HR) and vehicular-based signals will represent the drowsiness level of the driver more accurately, compared to only vehicular-signals."*
H2c: *"Combining the Heart Rate Variability (HRV) and vehicular-based signals will not represent the drowsiness level of the driver more accurately, compared to only vehicular-signals."*
H2d: *"Combining the Heart Rate (HR) and Respiration Rate (RR) will represent the drowsiness level of the driver more accurately, compared to only vehicular-signals."*
H2e: *"Combining the Heart Rate (HR), Heart Rate Variability (HRV) and Respiration Rate (RR) will represent the drowsiness level of the driver more accurately, compared to only vehicular-signals."*

Method

To answer the research question, this research makes use of an already existing dataset. The data was collected for Philips in 2011 and 2012 by TNO (Netherlands Organisation for Applied Scientific Research) and was provided to the author by Philips Research. An introduction to the dataset and the procedure used for gathering it is presented in the following section, followed by the pre-processing steps applied to the data and the statistical tests to be performed to answer the RQ and test the hypotheses.

Design

The experiment utilized a within-subject design. The goal of the initial experiment was to determine the feasibility of detecting drowsiness in subjects while driving, by analyzing trends of vital signs (for a full list of the vital signs recorded, see Table 1 under column "*Physiology*").

Participants

For the experiment, 27 participants were recruited. All participants were males, between 25 and 45 years of age and had a minimum of two years of driving experience with over 10,000 km/year. The experiment took place in a driving simulator and, thus, participants had to be tolerant to simulation sickness. Tolerance to simulation sickness was assessed during the initial stages of the experiment, while participants that reported. The exclusion criteria included the use of sleep medication and sleep-related disorders, the consumption of either more than 6 cups of coffee per day or 21 alcoholic beverages per week, to overcome possible withdrawal effects. Out of the total 27 participants, complete data is available for 19 participants. The remaining 8 participants did either not complete all sessions or their data was not correctly recorded and, thus, their data is not considered.

Apparatus & Measurements

In Table 1 an overview of the measurements recorded is provided. All the physiological measures were recorded using BioSemi equipment (developed by Biosemi, Inc., more information available at www.biosemi.com). The vehicular measures were obtained from the driving simulator and subjective measures were recorded via questionnaires that the participant filled in every five minutes during the experiment. Behavior (video) measures were recorded using a camera located on the dashboard and were manually labeled for yawns.

Physiology	**Vehicle**	**Behavior**	**Subjective**
EEG	Speed	Yawns	Karolinska Sleepiness Scale (KSS)
ECG	Acceleration		Visual Analogue Scale (VAS) for Relaxation
EMG	Lateral Position		
EoG	Steering Wheel position		
GSR	Lane ID		
Respiration	Pedal Position (brake/gas)		
	Time stamps of questionnaires		

Table 1
Complete list of the measurements/variables recorded during the experiment by TNO for Philips in 2011 and 2012.

Experimental Protocol

Before the experiment, participants were presented with a list of products containing caffeine and were instructed to refrain from their consumption on the day of the experiment to avoid interpretation problems it could cause on the physiological measurements. Additionally, they were further instructed to not consume any alcoholic product 24 hours prior to the experiment. On the day of the experiment the participant was welcomed and informed about the purpose of the experiment, and they were provided with the consent form. If the participant read and agreed with the text, they were are asked to sign the form in order to participate in the experiment. The experiment consisted of three sessions, for a total duration of 8.5 hours. Every participant, followed all three sessions on the same day. Let the participant indicate that they were not alert (e.g., due to not having a good night sleep), they were scheduled to participate on a different day.

The first session took place in the morning from 10 to 11 a.m. and served both as a baseline measurement and the test for simulation sickness tolerance. During this session, traffic conditions were indicated to be not monotonous to support the user to remain alert during these measurements and they were provided with instructions to maintain specific "target speeds", which changed throughout the duration of this session. This was done to keep the participants more engaged and alert. The second and third session lasted from 4 to 7:30 p.m. and 8 p.m. to 12 (midnight), respectively, and during these sessions, they did not have any further tasks as in the first session (i.e., they did not have to maintain a specific speed). However, the road simulated had low traffic conditions, in order to make the participants drowsy. After the second session, participants were provided with dinner, which was controlled for not containing substances that can affect the alertness level of the participants.
The KSS values (drowsiness scores) were filled in by the participant every 5 minutes and had a 30-second period to respond, otherwise a KSS value of 10 would be automatically registered, indicating that the participant was asleep. The score of 10 is outside the range of the Karolinska Sleepiness Scale and, thus, was assigned to be used as a flag value.

Pre-processing of Data & Feature Extraction

All the signals acquired from the BioSemi equipment were exported in a single *".bdf"* file for each session. Due to the presence of multiple signals not utilized in the present study, new files were generated capturing only the signals of interest (i.e., ECG and Respiration signals). To visualize the raw signals, the EDF Browser software was utilized (van Beelen, n.d.), and the signals were further processed using the custom PhysioData Toolbox, developed by Sjak-Shie (2022), in MatLab.

Window Selection. The interest of the research lies in the relation of the dependent variable (KSS) and the predicting variables. However, since the KSS scores were measured only in distinct moments, a modification of the data was required to assess associations between the vehicular and physiological data, which were continuously monitored, and the KSS. This was done by using reference windows to reflect characteristics of the signals prior to the registered KSS value. It was decided to select a 3-minute period to avoid uncertainty about potential stimulation the participant received due to the way the KSS questionnaires were presented and also the lack of information of when the questionnaire was filled out (i.e., the window of reference ending the moment the KSS questionnaire was presented, not filled out). The length of these windows would suffice to have enough data to be able to calculate a HR, RR and HRV representative of the period prior to the KSS, while not including potential effects due to the presentation of the previous KSS questionnaires.

Furthermore, as the participants were engaged in the baseline session in a different way than in the other sessions (i.e., by not only driving and filling out the KSS questionnaires, but also to maintaining specific speeds), it was decided to not include the data of the first session in the analyses.

Physiological Features. To answer the research questions and hypotheses, the extraction of specific features from the physiological signals was required. With regards to the ECG signal, the Heart Rate (HR) and Heart Rate Variability (HRV), by means of Root mean square of successive RR interval differences (RMSSD), were calculated and from the respiratory signal, the Respiration Rate (per minute) was extracted. The calculations were performed using the PhysioData software. The low-pass and high-pass filters were manually adjusted to depict the patterns present in the raw signals of both ECG and Respiration. The final values for the ECG signal were 0.5Hz for the High-pass and 40Hz for the Low-pass, while the values used for the Respiration were 0.05Hz for the High-pass and 40Hz for the Low-pass filter.

Vehicular Features. As far as the vehicular data are considered, four features were calculated. The first feature is the Average Absolute deviation of Lane Position (AALP), which indicates the average distance of the vehicle from the center of the lane, independent of direction. Secondly, the Standard Deviation of Lane Position (SDLP) was calculated as a measure of disperse from the mean. The same features were calculated for the Steering Wheel Angle (SWA) and are namely: the Absolute Average SWA (AASWA) and the Standard Deviation of the SWA (SDSWA). The calculation of the vehicular measures were performed using a custom made script in MatLab.

Statistical Tests & Analysis

In order to answer the research question and address the hypotheses multiple models need to be created. An alpha level of .05 was used for all statistical tests and the statistical software used was SPSS (v.29.0). The models were created by means of a multilevel analysis, such that more accurate inferences can be drawn for both the group and individual level, allowing for clustering in the individual level. The initial models included combinations of only vehicular measures (i.e., AALP, SDLP, AASWA and SDSWA), utilizing a step-wise addition approach, in order to predict the KSS values. This means that as a starting point, the empty model was considered. In every iteration a vehicular variable was added to the model, and their added predictive power was assessed. To check for the presence of multicollinearity among the independent variables, the Variance Inflation Factors (VIF) were inspected. The approach proposed by Daoud (2018) was used as a metric of contribution to multicollinearity. According to Daoud (2018), VIF values greater than 5 indicate high correlation between the independent variables, while VIF values of 1 indicate there is no correlation. Values falling between the two are considered moderately correlated. In case of high correlations between the variables, the variable with the highest VIF value was removed from the model and the VIF values of the new model was computed again until there was no high correlation between the variables.

Once the vehicular models were in place, additional models involving the physiological measures (HR, HRV and RR) were created and processed by the same means and procedures. Initially the model involving Respiration Rate was considered and the explained variance of the model was used to address the first part of the research question.

To answer the second part of the research question, the best model that includes also vehicular- and physiological-based measures will be considered and compared to the best performing vehicular-only-based model. To test if one the models is better performing, the explained variance of each model was used. To address each hypothesis the following models were considered:
H1: The model only considering Respiration Rate as a predictor for KSS.
H2a: The model involving Respiration Rate and vehicular-based signals, compared to the best performing vehicular-based model.

H2b: The model involving Heart Rate and vehicular-based signals, compared to the best performing vehicular-based model.
H2c: The model involving Heart Rate Variability (RMSSD) and vehicular-based signals, compared to the best performing vehicular-based model.
H2d: The model involving Respiration Rate, Heart Rate and vehicular-based signals, compared to the best performing vehicular-based model.
H2e: The model involving Respiration Rate, Heart Rate and Heart Rate Variability (RMSSD) and vehicular-based signals, compared to the best performing vehicular-based model.

Lastly, since the aim of the experiment was to induce drowsiness to the participants, an additional mixed-model was generated to assess if that was the case for the participants. This model used drowsiness, as depicted in the KSS scores, and time (by means of the 5-minute interval enumeration) as the only predictor.

Results

In the experiment 19 male participants, aged 25 to 45 years old and with a minimum of two years of driving experience, participated in a 8.5-hour simulator driving task. The research question that this paper aims to answer is the extent to which Respiration Rate can reflect the drowsiness level of a driver, indicated by the KSS scores, and to what extent Heart Rate, Heart Rate Variability and Respiration Rate, can help to improve drowsiness predictions based on solely vehicular measures. To answer the research question, mixed models were created and the results are presented in this section.

To investigate how Respiration Rate (RR) can reflect KSS (H1), a mixed model with RR as the only predictor was constructed. This model yielded a significant effect for RR (β = -.21, *SE*= .018, $t(1584)$ = -11.6, *p-value* < .001). It showed that RR could explain 10.3% of the variance in KSS ($pseudo\text{-}R^2 = 0.103$).

To explore the second part of the research question, the vehicular models were explored. These are the models utilizing the Average Absolute deviation of Lane Position (AALP), the Standard Deviation of Lane Position (SDLP), the Absolute Average of the Steering Wheel Angle (AASWA) and the Standard Deviation of the Steeting Wheel Angle (AASWA). As a first step, the empty model was examined and the vehicular measurements were added in a step-wise manner. The model including all vehicular variables was able to explain the most variance in the KSS score. A collinearity test was then conducted to check for the presence of multicollinearity among the variables. For that purpose, the Variance Inlation Factors (VIF) were calculated. The results of the test showed that the AALP had a VIF of 11.45, indicating a high correlation between the variables, as suggested by Daoud (2018), and, thus, was excluded to avoid multicollinearity effects (e.g., unstable estimates). Testing for multicollinearity after the removal of the aforementioned variable indicated that there is no high correlation among the remaining variables (i.e., all VIFs were below 1.87). As a result, the final vehicular model included the SDLP, the SDSWA and the AASWA.

The best overall performing vehicular model was found to be the model involving all three variables with 7.1% of the variance of KSS being explained by the model. The specifics of the model are shown in Table 2. Note that even though the Standard Deviation of SWA (SDSWA) did not seem to have a significant effect, the exclusion of it slightly decreased the model performance by means of explained variance (7.1% against 6.7%).

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	5.491	.247	20.689	22.275	<.001	4.978	6.004
SDLP	.522	.051	1618.689	10.233	<.001	.422	.621
AASWA	-29.541	7.453	1620	-3.963	<.001	-44.160	-14.921
SDSWA	3.503	1.370	1618.074	2.558	.011	.817	6.190

a. Dependent Variable: KSS.

Table 2
Estimates of Fixed Effects for KSS. The best fitting vehicular model assessed.

To address the remaining hypotheses, five different models were created and compared to the best overall performing vehicular-based model, by means of the variance the model could explain (*pseudo-*R^2). The results are presented below:

In Table 3 the coefficients for the Respiration Rate (RR) and vehicular-based model are presented. The specific model indicated an explained variance of 17.4% (*pseudo-*$R^2 = 0.174$).

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	8.276	.340	80.933	24.366	<.001	7.600	8.952
RR	-.203	.018	1593.474	-11.459	<.001	-.238	-.168
SDLP	.499	.049	1618.736	10.181	<.001	.403	.596
AASWA	-28.879	7.169	1619	-4.028	<.001	-42.941	-14.816
SDSWA	3.131	1.318	1617.428	2.376	.018	.546	5.716

a. Dependent Variable: KSS.

Table 3
Estimates of Fixed Effects for KSS. The model including Respiration Rate and vehicular features.

In Table 4 the coefficients for the Heart Rate (HR) and the vehicular measures are shown, the overall model could explain 8.9% of the variance in KSS (*pseudo-*$R^2 = 0.89$).

Table 5 addresses H2c and namely looks into if the introduction of RMSSD in the vehicular-only model. This model had a *pseudo-*R^2 of .071 and RMSSD did not have significant effect on the KSS score, $\beta = .001$, $t(1435.35) = 1.221$, $p = .222$, indicating it was not a significant predictor for drowsiness (KSS).

To address hypotheses H2d and H2e, two additional models were created. The model utilizing Heart Rate and Respiration Rate, on top of the vehicular data, was able to explain 18.9% of the variance in KSS. All of the predictor variables showed a significant effect on the KSS score ($p < .05$), (for a more detailed overview see Table 6).

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	6.488	.367	84.245	17.684	<.001	5.759	7.218
HR	-.011	.003	1361.180	-3.693	<.001	-.016	-.005
SDLP	.501	.051	1618.711	9.824	<.001	.401	.602
AASWA	-30.388	7.426	1619	-4.092	<.001	-44.954	-15.821
SDSWA	3.644	1.364	1618.056	2.671	.008	.968	6.321

a. Dependent Variable: KSS.

Table 4
Estimates of Fixed Effects for KSS. The model including Heart Rate and vehicular features.

Estimates of Fixed Effects[a]

Parameter	Estimate	Std. Error	df	t	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Intercept	5.362	.272	28.215	19.711	<.001	4.805	5.918
RMSSD	.001	.001	1435.349	1.221	.222	.000	.002
SDLP	.521	.051	1618.395	10.231	<.001	.421	.621
AASWA	-29.325	7.453	1619	-3.935	<.001	-43.943	-14.707
SDSWA	3.456	1.370	1618.069	2.524	.012	.770	6.142

a. Dependent Variable: KSS.

Table 5
Estimates of Fixed Effects for KSS. The model including Heart Rate Variability (RMSSD) and vehicular features. RMSSD was not found to have a significant effect.

Lastly, to test H2b the HRV feature RMSSD was added to the model and led to a *pseudo-*R^2 of .071, where RMSSD did not have a significant effect on the model.

The overall best performing model, by means of explained variance, was the model involving Heart Rate, Respiration Rate, SDLP, the SDSWA and AASWA with 18.9% variance being explained. To show how the predictions of the model and the actual drowsiness (KSS) values are related, two figures are presented. Figures 1 & 2 are provided and show how KSS values and predicted drowsiness are related and how they differ throughout the duration of the experiment for two participants. Figure 1 aims to depict an example of where the model performs poorly, while Figure 2 serves as an example of a participant where the model predicts drowsiness more closely to the actually KSS scores of the participant.

To assess whether drowsiness was, indeed, induced in the experiment an additional mixed-model was generated. The output of the analysis indicated that time had a significant effect on the KSS score ($p < .001$) and that effect showed a positive contribution.

Estimates of Fixed Effects[a]

						95% Confidence Interval	
Parameter	Estimate	Std. Error	df	t	Sig.	Lower Bound	Upper Bound
Intercept	9.291	.428	170.117	21.705	<.001	8.446	10.136
RR	-.204	.018	1595.325	-11.529	<.001	-.238	-.169
HR	-.011	.003	1358.533	-3.891	<.001	-.016	-.005
SDLP	.479	.049	1618	9.755	<.001	.383	.575
AASWA	-29.754	7.140	1618	-4.167	<.001	-43.759	-15.750
SDSWA	3.275	1.312	1617.496	2.496	.013	.702	5.849

a. Dependent Variable: KSS.

Table 6
Estimates of Fixed Effects for KSS. The model including Heart Rate, Respiration Rate and vehicular features.

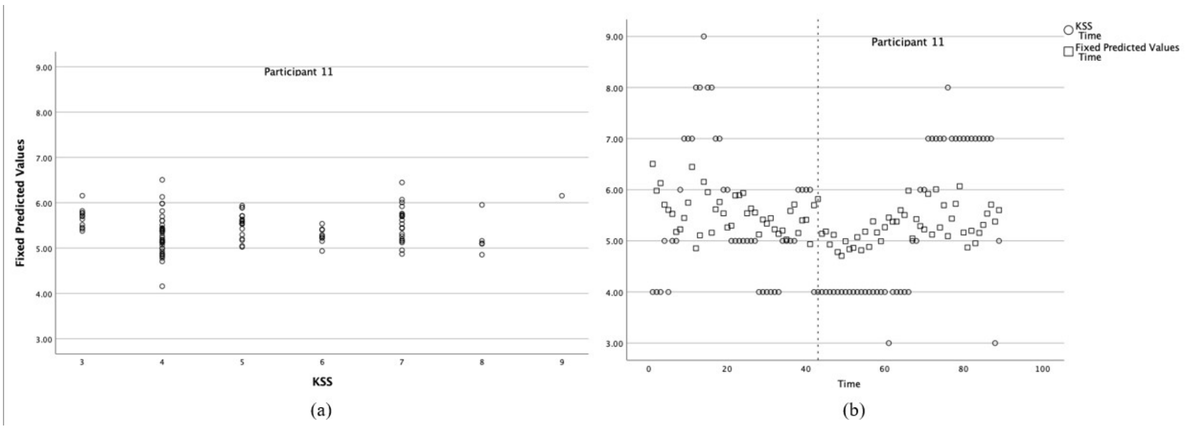

Figure 1. Two Figures (a) and (b) to illustrate how the model performs for participant 11. Figure 1(a), left, shows the predicted values and the actual KSS plotted against each other. Figure 1(b), shows the KSS values (denoted as a circle) and the predicted values (shown as squares) plotted against time. The vertical dashed line indicates the last measurement before the break. This serves as an example of how bad the model's performance can be, as the predictions mostly lie between the range 5 to 6, although the KSS values have a wider range.

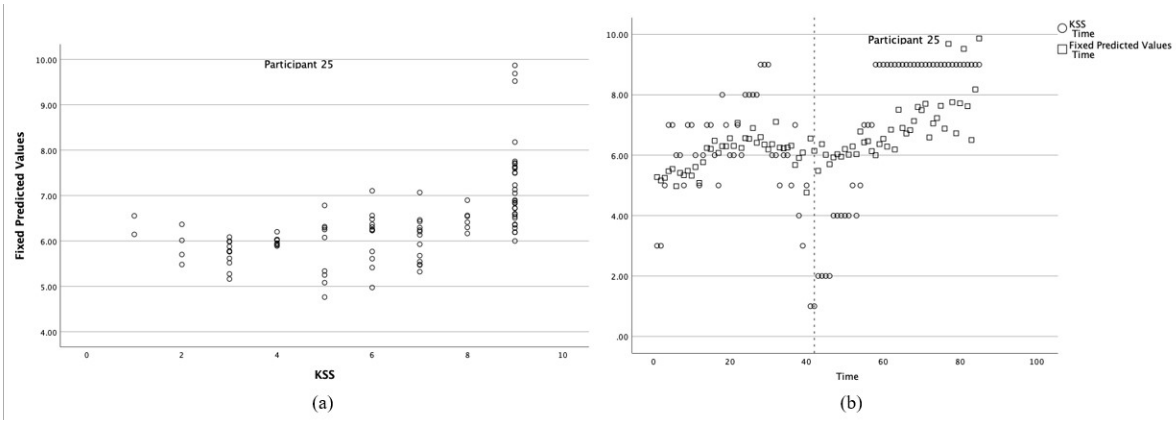

Figure 2. Two Figures (a), and (b) to illustrate how the model performs for participant 25. Figure 2(a), left, shows the predicted values and the actual KSS plotted against each other. Figure 2(b), shows the KSS values (circles) and the predicted drowsiness values (squares) plotted against time. The vertical dashed line indicates the last measurement before the break. The models seems to perform better for this participant, since the predicted values increase towards the end of and a small dip is observed around the break time.

Discussion

The main research question that this study aims to answer is the extent to which respiratory signal/patterns can reflect the drowsiness level of drivers. Based on the data, the feature Respiration Rate can reflect 10.3% of the variation of KSS and has a significant contribution to the vehicle-based prediction of driver drowsiness. This indicates that a decrease in Respiration Rate was associated with an increased self-reported drowsiness, meaning that a slow breathing pattern might be a sign of high drowsiness. This is in line with previous findings (Kiashari et al., 2018; Warwick et al., 2015).

To address the second part of the research question, the hypotheses will first be considered. All of the hypotheses concern the comparison between the explained variance by the physiological indicators above and beyond the variance explained by the extracted vehicle-based. Therefore, we established the best performing vehicular model (from the available vehicular features) as a baseline to which the models with physiological parameters are compared. Namely, this model involved the SDLP, AASWA and SDSWA and could explain 7.1% of the variance in KSS and will be referred to as the base model in this report. It is worth noting that the effects of SDLP and AASWA on the KSS score are in line with previous findings in the literature. SDLP has been indicated to increase for higher drowsiness scores and be a consistent predictor of drowsiness(Ingre et al., 2006) , while the decrease of Steering Wheel Movements (measured in AASWA) has also been reported to be associated with an increase in drowsiness (Feng et al., 2009). For the estimates of the effects see Table 2.

Hypothesis 1 stated that a hybrid-approach utilizing the Respiration Rate and vehicular-based features, will lead to a better performing model compared to the base model. As mentioned in the previous section, adding respiration rate to the base model could account for 17.4% of the variance, indicating that H1a was confirmed in the dataset. To address H2a, the model to be considered utilizes Heart Rate on top of the basic model. This model was able to explain 8.9% of the variance in KSS, indicating that H2a was also confirmed. However, the effect of HR on KSS score was found to be negative, meaning that the increase in HR was associated with lower drowsiness. This is not in line with the suggestion of Warwick et al. (2015), which indicate that Heart Rate increases while transiting to a drowsy state. However, it does seem to be in line with the findings of Jo et al. (2019). In their work they indicate that on average there is an increase of HR when a person drives compared to the daytime average, but that HR decreases from normal driving to drowsy driving.

Based on past research performed by van den Berg et al. (2005) and Warwick et al. (2015), where they indicate that HRV features do not show a significant change between awake and sleep states and, thus, should not be used as an indicator of sleepiness, it was hypothesized that adding RMSSD as a predictor to the base model will not lead to more explained variance. Adding the RMSSD as a predictor did not lead to the increase in the variance explained by the model and, thus, H2c was also validated. The validation of hypothesis 2c, seems to be in line with previous findings.

H2d and H2e, were both proved from the dataset. Namely, the models additionally utilizing Heart Rate and Respiration Rate, for H2d, and Heart Rate, Respiration Rate and RMSSD, for H2e, were both confirmed, with 18.9% and 17.8% variance explained by the models, respectively. Hence, the current findings suggest that involving Heart and Respiration Rate does, indeed, improve the base vehicular model and can explain 18.9% of the variance. To the author's knowledge, there is no hybrid application that combines only these measures. This is: there is no approach that only combined HR, RR and vehicular measures, without the inclusion of other physiological or behavioral measures. It is unclear to the author why this is the case, but based on the literature accessed, it seems that Respiration Rate is often not directly measured but its associated features are computed from different signals (e.g., ECG signal). However, models that do include them had an accuracy of 63% (Ingre et al., 2006).

It is important to note that even though 18.9% of the variance being explained by the model, and the predictors showing a significant effect, the importance of the results vary per application and domain. From an academic perspective, this indicates that future research may be required to look into whether such an effect, of Heart Rate and Respiration Rate in combination with some vehicular-based features having a significant effect on perceived drowsiness (KSS), is also present in different datasets and that this kind of data may be further utilized in more advanced models (i.e., Machine Learning or Deep Learning, to allow for more complicated interactions between the variables). However, for industry research, where the creation of a product is the primary goal, a model that could predict only 18.9% of the times correctly seems to be unreliable. According to their review of the existing literature on Drowsiness Detection techniques Hussein et al. (2021) physiology-based models have been studied, but their results by means of performance are not explicitly reported. This appears to also be the case for hybrid-approaches. A consequence of the absence of such information is that it cannot be indicate how well the models proposed in this research perform compared to the already studied models.

Although specific vehicular measures (i.e., SDLP) have been indicated to be consistent in predicting drowsiness (Liu, Hosking, & Lenné, 2009), the base model of this research was only able to explain a small amount of variance (7.1%). While 18.9% may be a low amount of variance for an overall model, it is still more variance than the base model could explain. Based on the idea that HR and RR were able to increase the predictive power of the vehicular system, the author believes that further investigation on such systems (i.e., based on HR, RR and vehicular signals) can still be fruitful and provide relevant to drowsy driving insights. Potential explanations of the low variances explained by both the base and best performing model of this research might may originate in the data analyzed. However, as mentioned in previous sections, there are a lot of uncertainties about the experiment and how the data was collected, and, thus, the author can only speculate about things that could have gone sideways (e.g., potential sensor malfunctions). Another potential explanation may be the way participants used the KSS scoring system. The usage of subjective measures can already cause discrepancies between participants, as it represents the perceived level of drowsiness. This being said, it is possible that participants indicated a high score early in the experiment, but there was no way to retract that score and, such, the only logical way to indicate they were more drowsy would be to further increase the reported score. This could, potentially, explain why (most) predictors had an effect in line with previous findings by means of the direction of their contribution in the prediction (i.e., positive or negative), but the overall model could not explain the variance in a high degree.

Limitations & Future Work

The analysis performed in the present research makes use of an already existing dataset that was gathered over a decade ago. However, this does not deem the dataset irrelevant. During this time a lot of new technologies, with regards to physiological measurement devices, have been available both for research and public use. Wireless and non-intrusive devices can now be used to gather physiological data from drivers (e.g., cameras), making some of the initial drawbacks and concerns for measuring devices disappear or have less effect on the measured variables (Solaz et al., 2016). This gives a lot of opportunities for the specific field to explore different approaches and develop different systems, but does not change the fact that these new technologies measure or try to approximate the same measurements that were directly recorded in the dataset used for the purpose of this research.

Nonetheless, the results of this research should be handled with caution, since there are multiple factors that could have affected the measurements and could not be accounted for, due to the absence of specific information. Initially, the way that participants were prompted to fill out the KSS questionnaire is unknown. This is of great importance, since the questionnaire was repetitively presented to the participants with 5-minute intervals. Let the questionnaire be accompanied by a sound-indicator or a flashing light, this could have caused the participant to be more alert while assessing their drowsiness level, as compared to the minutes before the presentation. The physiological measurements, however, still reflect drowsiness before the presentation of the questionnaire. As a result, the participants could have reported their momentary feeling of drowsiness, which was decreased due to the presentation method (i.e., either flashing light or sound), creating a discrepancy between the dependent and independent variables. Secondly, it is also unknown whether blue-light filters were placed on the simulator to account for the reduced presence of blue light in more realistic driving scenarios during night. The absence of such filters might also have led to a reduced level of drowsiness, affecting the reported KSS values. This could, in turn, have an effect on the overall model performance, assuming that high values of KSS were less present in the data, as compared to a real-life driving scenario. It is suggested that for future research involving Drowsiness Detection Systems, such factors are accounted for by utilizing a different drowsiness assessment tool, such as the use of objective criteria/behaviors and trained observers, as proposed by Wierwille and Ellsworth (1994), which would not intrude and, thus, not interact the flow of the participants.

The experiment during which the data was gathered involved three sessions with a total duration of 8.5 hours in a single day. This does not seem as a reasonable driving time in a single day for most individuals, but rather as a way to ensure the participants would feel drowsy by the end of the experimental day, which was found to also be the case. Additionally, in the experimental protocol it is mentioned that the participants were instructed to not eat any foods that could increase the production of melatonin or cortisol, hormones known to be associated with the regulation of sleep-wake cycles, or consume coffee. This might be useful to detect patterns and relations between physiological- and vehicular-based features, while transiting to a more drowsy state. However, for future research, it might be of interest to allow the participants to perform their daily activities without constraints, to generate data more representative of everyday driving scenarios and, potentially, increase the ecological validity of models for the detection of drowsiness in drivers.

In the current paper the decision based on which the best model is indicated is the use of the explained variance of each mixed-model generated (*pseudo*-R^2). While the direct comparison can be informative when comparing models that aim to predict the same outcome, it does not allow for the direct comparison in different contexts, nor about whether the improvement between two models is statistically significant. It is suggested that for future research a statistical test, such as the Likelihood Ratio Test (LRT), is used to check if the fit of one model is statistically better.

Bella (2013) in their work address another limitation of driving simulators and the research utilizing them, namely: drivers to not perceive any risk. This being said, behaviors observed in a simulator may not correspond to real world settings (i.e., driving an actual car). While this may not pose a problem while trying to identify associations between drowsiness and physiological variables, it may affect vehicular measures. As discussed in Liu et al. (2009), simulation experiment tend to represent monotonous streets (i.e., absence of strong turns) with low traffic conditions. It might be of interest for future research to introduce more complicated driving scenarios, or include further variables that may affect the driving behavior and the need for adjustments (i.e., side wind push), to be able to have more realistic simulations and, thus, potentially increase the ecological validity of the simulation findings. This need not be directly implemented for studies that aim to understand how drowsiness is reflected in different signals, but it should be considered for the development and test of systems that will be used in the real world.

Lastly, for the creation of drowsiness detection systems there are two steps that need to be made. Making a model that can predict drowsiness is the first step of the process. The aim of this study was to contribute to the knowledge of how well physiological measures (from HR, HRV and RR) and vehicular measures (SDLP, AALP, AASWA and SDSWA) can predict drowsiness by means of the KSS scores. The second step involves the prevention of an accident, by potentially providing an alert to the driver. In the current literature, there seems to be no study investigating when a driver should be considered "not capable of driving" and, thus, what the optimal moment to interfere is. It may be useful for future research to define a distinct point or range in the Karolinska Sleepiness Scale (or any other scale used to assess drowsiness), where a driver is considered not safe to drive. This will, ultimately, allow the computation of accuracy measures that have the same classification threshold (i.e., when the driver should be warned), even when different datasets or techniques are used, and the comparison based on accuracy of a model may be more sensible. For instance, in their work Worle, Metz, and Prill (2023) define a person too drowsy to drive when they report a KSS score of 8 twice in a row, but this is not consistent in the literature.

Based on the limitations faced in this study, an experimental set-up and methodology for future research can be found in the Appendix. This aims to serve as the methodology the author would have preferred in case they had the opportunity to collect the data on their own.

Conclusion

The aim of this study was to investigate to what extent physiological measures, such as Heart Rate, Respiration Rate and Heart Rate Variability (RMSSD) can potentially be used in combination with the currently used approaches to Driver Drowsiness Detection systems, which mainly involve measures obtained from the vehicle itself. In the literature, it is reported that vehicular-based systems may not detect changes in the driving behavior fast enough to prevent an accident and that the detection of some behaviors might be dependent on external factors to the drivers (i.e., appropriate road marking and lighting conditions). Physiological measures, on the other hand, are reported to be able to detect changes in early stages of the transition from an awake to a sleepy state, where more time to react is available and, thus, avoiding an accident might be more likely. Even though many researchers seem to acknowledge this, there are no systems, to the authors' knowledge, that utilize a hybrid approach including Heart Rate, Respiration Rate and Heart Rate Variability features in combination with vehicular-only based models to predict the drowsiness of drivers. However, combinations of the aforementioned predictors have been assessed in the literature (e.g., Heart Rate and Breathing Rate model by Warwick et al. (2015)). Based on the absence of such an approach, models combining these measurements are created and assessed. In conclusion, both Respiration Rate and Heart Rate can be viable predictors for driver drowsiness, while Heart Rate Variability (RMSSD) does not. Based on the results of the research, there is proof that a hybrid-approach, involving both Heart Rate and Respiration Rate on top of vehicular measures, might be able to improve the drowsiness prediction over the currently vehicular-only approaches.

Acknowledgment

First and foremost, I have to thank Joyce Westerink for her constant guidance, support, input and time invested throughout my graduation project, were she acted as both my supervisor from the Human-Technology Interaction group at the Technical University of Eindhoven and Philips Research. Secondly, I would like to thank Karin Smolders from the Human-Technology Interaction group at the Technical University of Eindhoven, the second supervisor of this thesis, for her contribution as well as her comments and input that led to the execution of this thesis project. Finally, I would like to thank Philips Research for giving me the opportunity, as well as the dataset and resources, to perform this research during my internship.

References

Albadawi, Y., Takruri, M., & Awad, M. (2022, 3). A Review of Recent Developments in Driver Drowsiness Detection Systems. *Sensors*, *22*(5), 2069. Retrieved from `https://www.mdpi.com/1424-8220/22/5/2069` doi: 10.3390/s22052069

Artaud, P., Planque, S., Lavergne, C., Cara, H., De Lepine, P., Tarriere, C., & Guenguen, B. (1994). An on-board system for detecting lapses of alertness in car driving. In *Proceedings of the 14th international technical conference on enhanced safety of vehicles.*

Bajaj, J. S., Kumar, N., & Kaushal, R. K. (2021). Comparative Study to Detect Driver Drowsiness. In *2021 international conference on advance computing and innovative technologies in engineering, icacite 2021* (Vol. 7, pp. 678–683). doi: 10.1109/ICACITE51222.2021.9404761

Balasubramanian, V., & Adalarasu, K. (2007). EMG-based analysis of change in muscle activity during simulated driving. *Journal of Bodywork and Movement Therapies*, *11*(2). doi: 10.1016/j.jbmt.2006.12.005

Bella, F. (2013). Driver perception of roadside configurations on two-lane rural roads: Effects on speed and lateral placement. *Accident Analysis and Prevention*, *50*. doi: 10.1016/j.aap.2012.04.015

BioHarness 3. (n.d.). Retrieved from `https://www.zephyranywhere.com`

Chen, L. l., Zhang, A., & Lou, X. g. (2019). Cross-subject driver status detection from physiological signals based on hybrid feature selection and transfer learning. *Expert Systems with Applications*, *137*. doi: 10.1016/j.eswa.2019.02.005

Daoud, J. I. (2018). Multicollinearity and Regression Analysis. In *Journal of physics: Conference series* (Vol. 949). doi: 10.1088/1742-6596/949/1/012009

Doudou, M., Bouabdallah, A., & Berge-Cherfaoui, V. (2020). Driver Drowsiness Measurement Technologies: Current Research, Market Solutions, and Challenges. *International Journal of Intelligent Transportation Systems Research*, *18*(2), 297–319. doi: 10.1007/s13177-019-00199-w

Feng, R., Zhang, G., & Cheng, B. (2009). An on-board system for detecting driver drowsiness based on multi-sensor data fusion using dempster-shafer theory. In *Proceedings of the 2009 ieee international conference on networking, sensing and control, icnsc 2009.* doi: 10.1109/ICNSC.2009.4919399

Finelli, L. A., Baumann, H., Borbély, A. A., & Achermann, P. (2000). Dual electroencephalogram markers of human sleep homeostasis: Correlation between theta activity in waking and slow-wave activity in sleep. *Neuroscience*, *101*(3). doi: 10.1016/S0306-4522(00)00409-7

Gonçalves, J., & Bengler, K. (2015). Driver State Monitoring Systems– Transferable Knowledge Manual Driving to HAD. *Procedia Manufacturing*, *3*. doi: 10.1016/j.promfg.2015.07.845

Gwak, J., Hirao, A., & Shino, M. (2020). An investigation of early detection of driver drowsiness using ensemble machine learning based on hybrid sensing. *Applied Sciences (Switzerland)*, *10*(8). doi: 10.3390/APP10082890

Hoddes, E., Zarcone, V., Smythe, H., Phillips, R., & Dement, W. C. (1973). Quantification of Sleepiness: A New Approach. *Psychophysiology*, *10*(4). doi: 10.1111/j.1469-8986.1973.tb00801.x

Hussein, M. K., Salman, T. M., Miry, A. H., & Subhi, M. A. (2021). Driver Drowsiness Detection Techniques: A Survey. In *1st babylon international conference on* *information technology and science 2021, bicits 2021* (pp. 45–51). doi: 10.1109/BICITS51482.2021.9509912

Ingre, M., Åkerstedt, T., Peters, B., Anund, A., & Kecklund, G. (2006). Subjective sleepiness, simulated driving performance and blink duration: Examining individual differences. *Journal of Sleep Research*, *15*(1). doi: 10.1111/j.1365-2869.2006.00504.x

Jo, S. H., Kim, J. M., & Kim, D. K. (2019). Heart rate change while drowsy driving. *Journal of Korean Medical Science*, *34*(8). doi: 10.3346/jkms.2019.34.e56

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology*, *117*(7). doi: 10.1016/j.clinph.2006.03.011

Katsis, C. D., Ntouvas, N. E., Bafas, C. G., & Fotiadis, D. I. (2004). Assessment of muscle fatigue during driving using surface EMG. In *Proceedings of the iasted international conference on biomedical engineering.*

Khushaba, R. N., Kodagoda, S., Lal, S., & Dissanayake, G. (2011). Driver drowsiness classification using fuzzy wavelet-packet-based feature-extraction algorithm. *IEEE Transactions on Biomedical Engineering*, *58*(1). doi: 10.1109/TBME.2010.2077291

Kiashari, S. E. H., Nahvi, A., Homayounfard, A., & Bakhoda, H. (2018). Monitoring the Variation in Driver Respiration Rate from Wakefulness to Drowsiness: A Non-Intrusive Method for Drowsiness Detection Using Thermal Imaging. *Journal of Sleep Sciences*, *3*(2), 1–9.

Lagarde, D., & Batejat, D. (1994). Evaluation of drowsiness during prolonged sleep deprivation. *Neurophysiologie Clinique / Clinical Neurophysiology*, *24*(1). doi: 10.1016/S0987-7053(05)80403-4

Li, Z., Li, S. E., Li, R., Cheng, B., & Shi, J. (2017). Online detection of driver fatigue using steering wheel angles for real driving conditions. *Sensors (Switzerland)*, *17*(3). doi: 10.3390/s17030495

Liu, C. C., Hosking, S. G., & Lenné, M. G. (2009). Predicting driver drowsiness using vehicle measures: Recent insights and future challenges. *Journal of Safety Research*, *40*(4). doi: 10.1016/j.jsr.2009.04.005

Ngxande, M., Tapamo, J. R., & Burke, M. (2017). Driver drowsiness detection using behavioral measures and machine learning techniques: A review of state-of-art techniques. In *2017 pattern recognition association of south africa and robotics and mechatronics international conference, prasa-robmech 2017* (Vol. 2018-Janua). doi: 10.1109/RoboMech.2017.8261140

Nordbakke, S., & Sagberg, F. (2007). Sleepy at the wheel: Knowledge, symptoms and behaviour among car drivers. *Transportation Research Part F: Traffic Psychology and Behaviour*, *10*(1). doi: 10.1016/j.trf.2006.03.003

Pack, A. I., Pack, A. M., Rodgman, E., Cucchiara, A., Dinges, D. F., & Schwab, C. W. (1995). Characteristics of crashes attributed to the driver having fallen asleep. *Accident Analysis and Prevention*, *27*(6). doi: 10.1016/0001-4575(95)00034-8

Philip, P., Sagaspe, P., Moore, N., Taillard, J., Charles, A., Guilleminault, C., & Bioulac, B. (2005). Fatigue, sleep restriction and driving performance. *Accident Analysis and Prevention*, *37*(3). doi: 10.1016/j.aap.2004.07.007

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). *Detecting driver drowsiness based on sensors: A review* (Vol. 12) (No. 12). doi: 10.3390/s121216937

Satti, A. T., Kim, J., Yi, E., Cho, H. Y., & Cho, S. (2021). Microneedle array electrode-based wearable emg system for detection of driver drowsiness through steering wheel grip. *Sensors*, *21*(15). doi: 10.3390/s21155091

Shaffer, F., & Ginsberg, J. P. (2017). *An Overview of Heart Rate Variability Metrics and Norms* (Vol. 5). doi: 10.3389/fpubh.2017.00258

Sjak-Shie, E. E. (2022). *PhysioData Toolbox (Version 0.6.3).*

Slater, J. D. (2008). A definition of drowsiness: One purpose for sleep? *Medical Hypotheses*, *71*(5). doi: 10.1016/j.mehy.2008.05.035

Solaz, J., Laparra-Hernández, J., Bande, D., Rodríguez, N., Veleff, S., Gerpe, J., & Medina, E. (2016). Drowsiness Detection Based on the Analysis of Breathing Rate Obtained from Real-time Image Recognition. *Transportation Research Procedia*, *14*, 3867–3876. Retrieved from `http://dx.doi.org/10.1016/j.trpro.2016.05.472` doi: 10.1016/j.trpro.2016.05.472

Sunwoo, J. S., Hwangbo, Y., Kim, W. J., Chu, M. K., Yun, C. H., & Yang, K. I. (2017). Sleep characteristics associated with drowsy driving. *Sleep Medicine*, *40*. doi: 10.1016/j.sleep.2017.08.020

SWOV. (2019, 9). Fatigue. Retrieved from `https://swov.nl/en/fact-sheet/fatigue`

van Beelen, T. (n.d.). EDFbrowser. *[Computer software]. Retrieved from https: https://www.teuniz.net/edfbrowser/.*

van den Berg, J., Neely, G., Wiklund, U., & Landström, U. (2005). Heart rate variability during sedentary work and sleep in normal and sleep-deprived states. *Clinical Physiology and Functional Imaging*, *25*(1). doi: 10.1111/j.1475-097x.2004.00589.x

Warwick, B., Symons, N., Chen, X., & Xiong, K. (2015). Detecting driver drowsiness using wireless wearables. *Proceedings - 2015 IEEE 12th International Conference on Mobile Ad Hoc and Sensor Systems, MASS 2015*(October), 585–588. doi: 10.1109/MASS.2015.22

Wierwille, W. W., & Ellsworth, L. A. (1994). Evaluation of driver drowsiness by trained raters. *Accident Analysis and Prevention*, *26*(5), 571–581. doi: 10.1016/0001-4575(94)90019-1

Worle, J., Metz, B., & Prill, A. (2023). How to Induce Drowsiness When Testing Driver Drowsiness and Attention Warning (DDAW) Systems. *IEEE Transactions on Intelligent Transportation Systems*. doi: 10.1109/TITS.2023.3249755

Appendix
Suggested Method

In the following section, a framework for future research is suggested. It mainly serves as an alternative experiment that the author proposes in the scenario that they had the option to perform the data collection on their own. As such, when a change is proposed, it is accompanied with an explanation of why this choice may be benefitial compared to the initial experiment performed.

Design

It is suggested that the follow up experiment utilizes a within-subject design. The goal of the experiment remains the same and is, namely: to look into the feasibility of detecting drowsiness in subjects while driving, by analyzing trends of vital signs and vehicular signal patterns. To be able to address this goal, the experiment is again designed in such a way so that the participants get more drowsy throughout the duration of it.

Participants

Based on the review of Hussein et al. (2021), most experiments that involve physiological signals have about 15 participants, while there are also experiments that involved only 3 participants, but in multiple sessions. This being said, recruiting about 20 participants (10 males and 10 females) should give sufficient data, even in the case that some participants decide they do not want to continue with the experiment.

In the experiment performed by TNO for Philips, multiple inclusion criteria were in place. Initially, all participants were males. This choice was not justified, yet it seems reasonable to include both females and males in the experiment, as both genders are potential users of Drowsiness Detection Systems and it has been suggested in the literature that the magnitude of changes in some physiological signals vary for different genders (Warwick et al., 2015). Another criterion based upon participants were recruited was the minimum of two years of driving experience with over 10,000 km/year. To the author, there is no obvious reason for which such inclusion criteria should be in place and, thus, the possession of a driver's license would suffice, based on the idea that all drivers should be able to use such systems independent of their experience.

As far as the exclusion criteria are considered, drivers that make use of sleep medication or have sleep-related disorders were decided not to be included, and so did people than on average consume more than 6 cups of coffee or 21 alcoholic beverages per week. Since the aim of the research focuses on the relation between the physiological measures and drowsiness, it seems logical to try minimize extraneous factors that may have an effect on the variables involved in order to find the ground truth (i.e., the basis of how and whether drowsiness can be reflected in the physiological measures). As a result, the same exclusion criteria apply for the proposed experiment.

Apparatus & Measurements

The physiological measures will be recorded by means of wireless devices, such as BioHarness 3.0 which consists of a chest strap that can record, store and transmit vital data, such as ECG, Heart Rate, Respiration Rate and body orientation (*BioHarness 3*, n.d.). In the experiment performed by TNO for Philips, the BioSemi equipment was used, which consisted of multiple sensors placed over the body of the participant. Since the initial experiment was not performed for the purpose of this research, many of the signals recorded by the BioSemi equipment were not relevant for us and, thus, it would be preferred to only record the signals of interest, which can be obtained from the single chest strap BioHarness 3.0. The vehicular measures will be obtained from the driving simulator, as it was done in the experiment performed by TNO. As discussed in the Limitations & Future Work section of the report, having the participants fill out a questionnaire every five minutes may have interfered with the dependent variable and it was proposed to use objective criteria/behaviors and trained observers to take over this procedure. This being said, the drowsiness measurement in the proposed experiment will not be the KSS scores, but videos of the driver will be recorded and the scores will assigned by trained observers for each 5-minute segment of the drive (for more information see Wierwille and Ellsworth (1994)).

Experimental Protocol

Before the the experiment, participants will be presented with a list of products containing caffeine and are instructed to refrain from their consumption on the day of the experiment to avoid interpretation problems it could cause on the physiological measurements. Additionally, they will be further instructed to not consume any alcoholic product 24 hours prior to the experiment. On the day of the experiment the participant is welcomed and informed about the purpose of the experiment and they are provided with a consent form. If the participant reads and agrees with the text, they are asked to sign the form, in order to participate in the experiment.

According to an analysis performed by Pack et al. (1995), crashes attributed to drivers being sleepy mainly took place during two times of the day: the nigh time period (midnight to 7 a.m.) and during the mid-afternoon time (3 p.m.). Furthermore, in their review paper Liu et al. (2009) provide a summary of the methods used in research aiming to predict driver drowsiness from vehicular measures, including the drowsiness manipulations and the driving task. For the experiments not involving a sleep deprivation manipulation, it is observed that the total driving time is between 80 and 210 minutes, split into up to 6 sessions to test for various times of the day.

Based on the aforementioned information, it is proposed that the experiment consists of two sessions of 1.5 hours that take place on the same day. The first session will take place from 2 to 3:30 p.m., while the second session will take place from 11 to 12:30 p.m. to include the times that were reported by Pack et al. (1995). The traffic conditions in the simulation will be set to low and the street will represent a monotonous route (absence of strong turns), in order to induce drowsiness. Lastly, the lighting conditions should represent the conditions found in a real world setting. This includes both the room in which the simulator is place in, but also the simulated environment. To ensure no further blue-light is emitted from the screen, a blue-light filter or the monitor settings of the simulator will be applied.

Pre-processing of Data & Feature Extraction

Window Selection. The interest of the research lies in the relation of the dependent variable (drowsiness as assessed by the trained observer) and the predicting variables. Since the drowsiness scores will refer to distinct moments, a modification of the data is required to assess the associations between the vehicular and physiological, which will be continuously monitored, and the drowsiness score. Since there is no stimulation provided to the participant, a window representing the full 5-minute segment can be used.

Physiological Features. To answer the research question and the hypotheses, the extraction of specific features from the physiological signals will be required. Due to the use of BioHarness 3.0 equipment, the physiological features of HR and RR can be directly computed for each window, while the RMSSD will be calculated using MatLab (or software of choice).

Vehicular Features. As far as the vehicular data are considered, four features will be calculated. The first feature is the Average Absolute deviation of Lane Position (AALP), which indicates the average distance of the vehicle from the center of the lane, independent of direction. Secondly, the Standard Deviation of Lane Position (SDLP) will calculated as a measure of disperse from the mean. The same features will be calculated for the Steering Wheel Angle (SWA) and are namely: the Absolute Average SWA (AASWA) and the Standard Deviation of the SWA (SDSWA). The calculation of the vehicular measures will be performed using a custom made script in MatLab, which was also used to analyze the same data in the main report.

Statistical Tests & Analysis

The same statistical tests and analysis will be performed as discussed in the main Methods section, with the addition of an extra test: the Likelihood Ratio Test (LRT). The LRT will be used to determine whether a more complex model (i.e., including multiple predictors) provides a statistically better fit on the data than a simpler model.