Fast Detection and Classification of Drivers' Responses to Stressful Events and Cognitive Workload

Fabien Rogister[1(✉)], Marie-Anne Pungu Mwange[1], Luka Rukonić[1,2], Olivier Delbeke[1], and Richard Virlouvet[1]

[1] Connected and Sharing Solutions, AISIN Europe, Braine-L'Alleud, Belgium
fabien.rogister@aisin-europe.com
[2] Institute for Language and Communication, Université catholique de Louvain, Louvain-la-Neuve, Belgium

Abstract. We apply machine learning techniques to detect moments of stress and cognitive load during simulator driving experiences. The use of the electrical skin conductance, or more precisely the electrodermal activity (EDA), is particularly interesting for assessing drivers' states because it is easily measurable; it is also involuntary and uncontrollable. Detection of responses to external stimuli can be performed on a scale of seconds with an accuracy of 86%. Moreover, we observe that responses to stress events and cognitive efforts can be differentiated with an accuracy of 80% over sub-minute time intervals. We compare our results to others reported in the literature. Automatic and fast detection of responses to stressful events and high cognitive workload can be used to assess drivers' user experience (UX) and their interaction with their vehicle.

Keywords: Machine learning · Driver state recognition · User experience · Electrodermal activity

1 Introduction

While aiming to make the driving experience more enjoyable and safer, it is highly desirable to automatically detect, avoid or alleviate stressful or cognitively demanding situations. Continuous monitoring of driver state is necessary to reach this goal.

Many studies are based on the analysis of physiological signals such as respiration, temperature, electromyograms, electrocardiograms (ECGs), and electrodermal activity (EDA), i.e. the variation in skin conductance. Their objectives are similar: the detection of the driver's state and its automation using machine learning techniques. These techniques have been applied to physiological signals measured on drivers, both in real driving (see [1, 2] and references therein) and on a simulator (see [3, 4] and references therein) to detect reactions to stressful events with high accuracy. However, they combine several physiological signals and the classification between several stress states relies on the analysis of long signal intervals, thus preventing the detection of driver states on short time scales.

C. Stephanidis et al. (Eds.): HCII 2022, CCIS 1581, pp. 210–217, 2022.
https://doi.org/10.1007/978-3-031-06388-6_28

In this study carried out in a simulated environment, our objective was to automatically detect on the shortest possible time scale, the physiological responses of drivers to external stimuli, and to determine those which are induced either by stress events or by cognitive load. To this end, we first considered the characteristics extracted from the electrodermal activity and from signals derived from the electrocardiogram, namely the heart rate and its variability. We nevertheless realized that the analysis of the EDA alone allowed a rapid detection and classification of driver states with an accuracy of 80%.

In this paper, we briefly present our driving simulator and the experimental scenarios we used to train and test our models. We investigate qualitatively the physiological responses to stimuli of participants to our experiments. We examine their electrodermal activity, heart rate, and heart rate variability. We show that, in our opinion, these last two signals are not relevant for fast detection of the responses to stress and cognitive tasks. We present the two successive steps of our method to automatically classify drivers' states.

2 Simulator and Electrophysiological Signals Measurement

The simulator is based on a customized version of CARLA [5], an open-source simulation environment based on Unreal engine. The setup consists of three large 50-inch-curved screens, an adaptable car seat and a Fanatec® set of a steering wheel, a gear shifter and pedals. Two cameras and a microphone were also used to record the experiment. A small auxiliary screen displayed a simplified navigation map.

Drivers' physiological signals were recorded at 2000 samples per second with a BIOPAC® MP36R connected to participants with adhesive electrodes. Two electrodes were positioned on the sole of the left foot for the EDA measurement; ECG electrodes were positioned close to the left medial malleolus and on the right wrist. The sole of the left foot for the EDA measurement was chosen to avoid motion artifacts, the simulator replicating an automatic transmission car.

3 Detection of Physiological Responses to External Stimuli

We have designed a first scenario in which we alternate periods of normal driving with stressful situations and situations requiring cognitive effort. A preliminary analysis of how these signals were varying with the succession of events allowed us to understand that the signals derived from the ECG, the heart rate and its variability, were not relevant on short time scales. We, therefore, decided to use EDA alone. In this section, we briefly present this scenario and our preliminary observations. We then explain the algorithmic steps we have implemented for the automatic detection of skin conductance responses to stimuli.

3.1 First Scenario and Preliminary Observations

The first scenario was completed by seven participants. They started in a town and after a few hundred meters they continued on a highway with a speed of 90 kph.

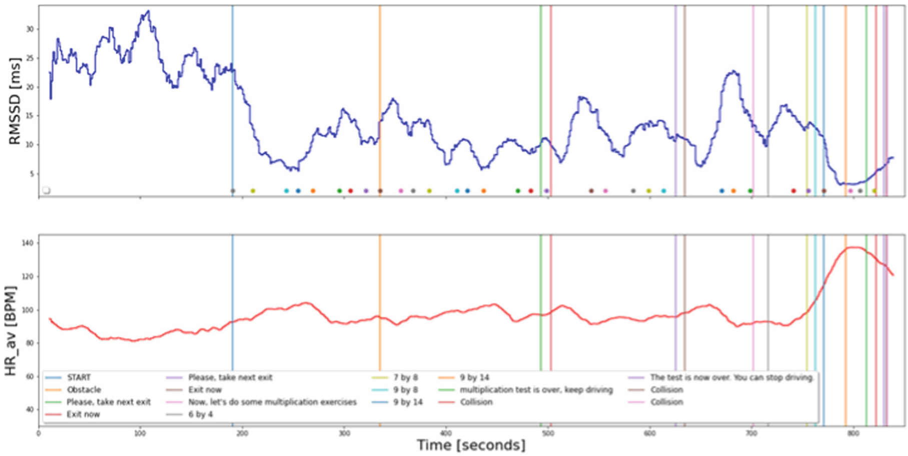

Fig. 1. RMSSD and heart rate (HR) of participant 7 enrolled in the first experiment. RMSSD is calculated over a 20 s window sliding by steps of one second. HR is averaged over 20 s. The vertical lines correspond to events (obstacle avoidance, interaction with the vocal assistant, mathematical exercises, etc.). The small dots correspond to turns on the highway. During the cognitive exercises, we observe an important increase in the heart rate while its variability decreases.

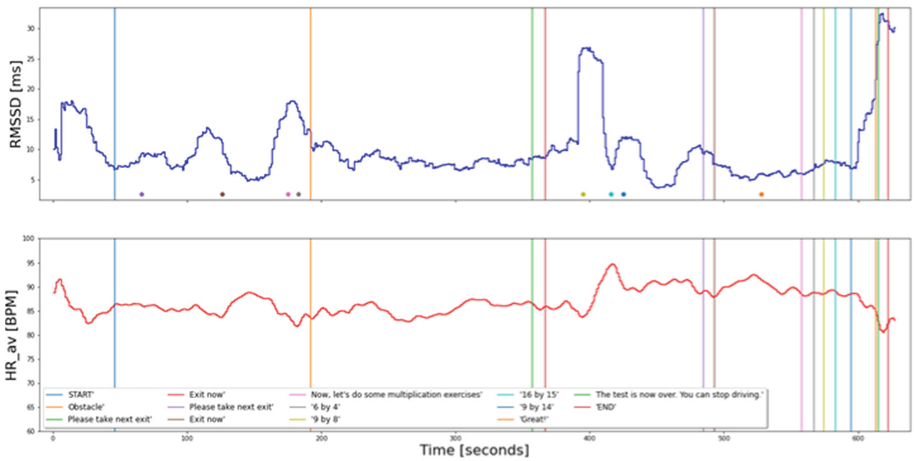

Fig. 2. RMSSD and heart rate (HR) of participant 2 enrolled in the first experiment. There is no observable increase in the heart rate and decrease in its variability concomitant with stressful events and cognitive tasks.

During the drive, which lasted approximately 12 min, they had to avoid an obstacle, then they were instructed by a voice assistant to take an exit and then come back on the highway. Later, they were instructed by the voice assistant to take another exit and then come back again on the highway. They were submitted to a few multiplication exercises in the last part of the scenario.

A qualitative analysis of heart rate and its variability, which are both derived from the ECG, and the EDA signal of the seven participants in this first experiment has been carried out. It is well known that heart rate increases, and its variability decreases in presence of stressful stimuli, but we have observed this trend in few situations only (Fig. 1 provides an example); this behavior is indeed not observed in five out of seven recordings (Fig. 2 provides an example where this trend is not observed). This is a noticeable difference with other studies based on ECG derived signals [1–4]. We explain this discrepancy by the fact that, in our experiments, stressful stimuli and periods of relative calm alternate on short timescales while, to the best of our knowledge, most studies consider a succession of longer sequences with very distinctive levels of stress.

EDA refers to variation in skin conductivity. It is linked to the sympathetic nervous system and is an indication of physiological and psychological arousal. It can be used as an indicator of cognitive and emotional states. Figure 3 presents an EDA signal, which consists of a succession of peaks, the skin conductance responses (SCRs). These responses can be spontaneous, but most are generated by stimuli. Both are unconscious and uncontrollable. The figure shows that most SCRs are following well-defined stressful events (such as sharp turns on the road, the avoidance of an obstacle or the interaction with the voice assistant) or are generated in a repetitive way during the cognitive tasks at the end of the scenario. We, therefore, focus in the following on EDA signals only for stress and cognitive load detection without paying more attention to ECGs.

3.2 First Classification Algorithm

Artifacts in raw EDA were removed with a median filter with a local window of one second. As the process is time-consuming, the signal was first decimated by a factor of 10 after applying an order 8 Chebyshev type I anti-aliasing filter. After removing artifacts, the signal was resampled using Fourier method. A clean EDA and its fast-changing component (namely the phasic component) were then calculated using the Python NeuroKit2 library [6]. Phasic was obtained by high-pass filtering the EDA. Figure 3 displays an EDA signal and its phasic component. The energy of the phasic signal, its mean absolute value, its mean absolute derivative, its maximum absolute derivative, its mean derivative and its mean absolute derivative are calculated for every 5-s-long intervals. The last two features are observed to be the most important to achieve an accurate classification. Each of the 5-s-long intervals are also labeled by one of the researchers as 0' in the absence of SCR, or conversely as 1' in the presence of response. For this task, he is assisted by the BIOPAC Acqknowledge software and functions from NeuroKit2 library. It must be pointed out that a small percentage of responses detected by both tools is questionable and we think this is an important source of errors in the subsequent supervised classification. Five different classifiers have been trained and tested by using the leave-one-out validation method: a k-Nearest Neighbors (k-NN) classifier, a Random Forest (FR) classifier, a Support Vector Machine (SVM) classifier with linear kernel, a Support Vector Machine classifier with a Radial Basis Function (RBF), and a four-layer artificial neural network (ANN) classifier. Accuracy, sensitivity and specificity obtained after hyper-parameter tuning are of the same order, SVM with RBF and ANN performing slightly better. In the case of the ANN, they are 86%, 84%

and 87%, respectively. Figure 4 shows the result of the first classification, where intervals without any SCR are in green, while intervals with SCRs are in grey.

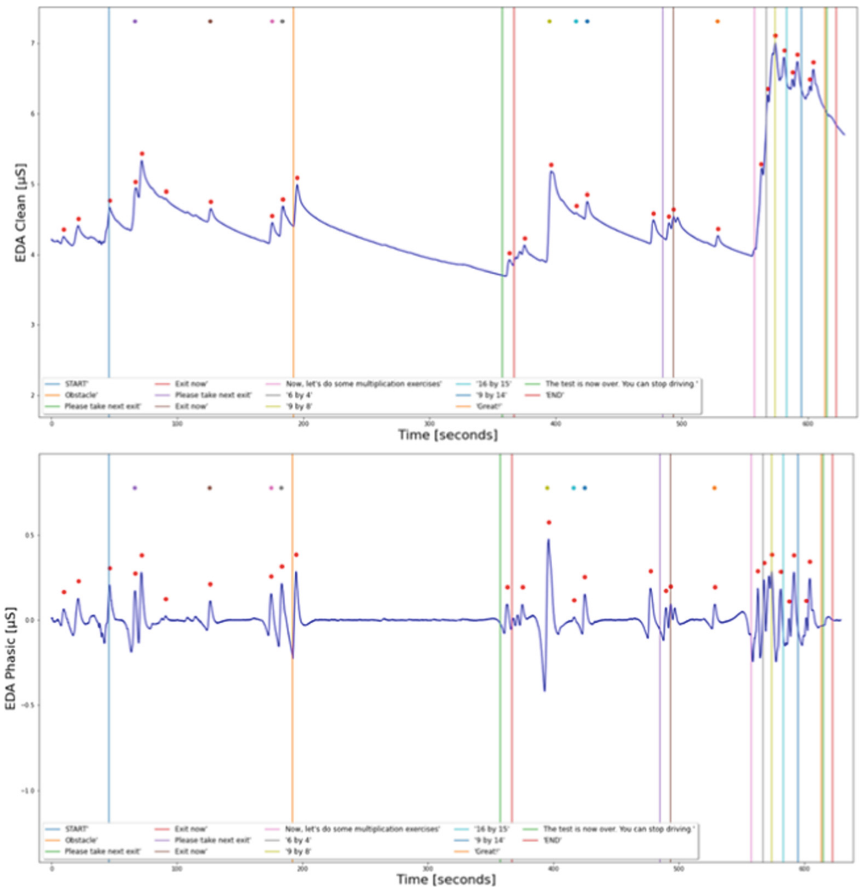

Fig. 3. Clean EDA and its phasic component of participant 2 enrolled in the first experiment. Red dots correspond to SCR peaks. The vertical lines and the other dots have same meaning as in Fig. 1. (Color figure online)

4 Classification of Physiological Responses to Stressful Events and Cognitive Tasks

The time duration devoted to the cognitive effort is very short in the first scenario and data collected were not sufficient to train and test a second classifier whose task would be to separate responses to cognitive effort from those relating to stressful events. Data collected during two other experiments have been used for this purpose. We describe below the scenarios, then the method used for the second classification process.

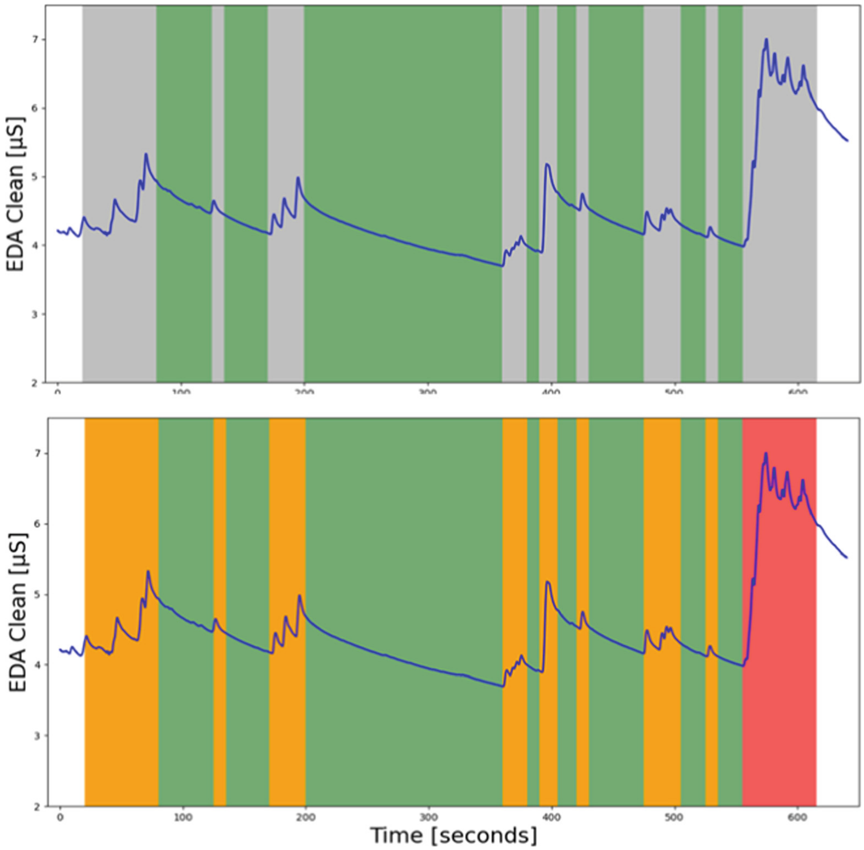

Fig. 4. Top: Result of the first classification. Intervals classified as without SCRs are in green, intervals classified as with SCRs are in grey. Bottom: Result of the second classification. Intervals corresponding to responses to stressful events are yellow. Intervals corresponding to responses to cognitive tasks are in red. (Color figure online)

4.1 Second and Third Scenarios

In the second scenario, participants were sitting in a real stationary car. They were instructed to first think about a song, then sing it in front of an experimenter while being video recorded. After relaxing, they had to add the number 3333 to seemingly randomly generated 4-digit numbers. After a second relaxing time, they achieved a "Stroop color-naming task" [7]. 7 participants took part in this experiment.

In the third scenario, the participants were driving on a highway, sharing the road with other cars controlled by the simulator. They were successively submitted to an "auditory Stroop task" [7], they had to count backwards by steps of 3 then 7, and finally they had to spell words backwards.

4.2 Second Classification Algorithm

The 5 s long intervals previously labelled by the first classifier as 0' but surrounded by intervals labeled as 1' (thus with SCR peaks) are relabeled as 1'. All the neighboring intervals with the same label are then coalesced in longer intervals. The new intervals that are without any detected SCRs are relabeled as 0 and remain unchanged until the end of the process. All the other new intervals, which include at least one SCR, are relabeled as $1''$.

The second classification process aims to separate these intervals labeled as $1''$ between those corresponding to responses to stressful events from those corresponding to responses to cognitive load. These intervals will be ultimately labelled either as 1 or 2, respectively. Several sets of features have been investigated. The most relevant set of features for the classification of intervals of the same duration is the number of SCRs in the intervals, the sum of SCR peak amplitudes measured from the onset of the responses, and the ratio between the number of SCRs and the interval duration.

The training and testing of a Random Forest classifier were made with intervals clearly identified as either with responses to stressful events (label 1) or with responses to cognitive tasks (label 2). The accuracy of this binary classification is 80% for 40 s long intervals. It increases with their duration. By contrast, it rapidly decreases for shorter intervals. An example of this second classification is presented in Fig. 4.

5 Discussion

In this section, we compare the results of the two classification processes described above with the results of some other studies [1–4]; it is nevertheless necessary to be aware of the difficulty of making fair comparisons as the methodologies differ. Some works are indeed based on experiments with simulators, others use data recorded in real cars. All use resting periods as a baseline measurement, contrary to our work.

We can compare the results obtained with our first classifier with a study aiming at a binary classification between stress and no stress [1]. This study is based on scenarios alternating relatively short periods of driving with and without obstacles to avoid. There, the best accuracy is 88% with Long Short-Term Memory networks with 15 s long intervals. This is overall comparable to the results of our first classifier.

The studies aiming at the classification of different levels of stress (namely absence of stress, low stress, high stress) consider long, uninterrupted, and very distinctive periods of driving (city streets, highways) in addition to resting periods [2, 4]. In [2], an overall accuracy of 97% is obtained but with 5-min intervals and features extracted from multiple physiological signals (skin conductance, respiration, heart rate, electromyogram). Only features extracted from skin conductance, heart rate and respiration on 100 s intervals are considered in [4]. The overall accuracy for the three states classification is around 90% but it must be pointed out that this value is pulled up by the nearly 100% correct detection of the resting periods. Therefore, the binary classification accuracy between low stress and high stress is likely close to 80%. That study shares the same limitations as those of [2] as it uses the same database.

6 Conclusion

We have not only shown that it is possible to reliably detect physiological responses to stressful events and cognitive tasks but also that it is possible to separate them over sub-minute time intervals.

The tool we have developed here can allow the improvement of drivers' experience as we can monitor their emotional state. We already applied it to assess the physiological adaptation of drivers to our simulator [8]. This research takes place in a project that deals with semi-automated cars.

Acknowledgement. We acknowledge the support by the project VIADUCT under the reference 7982 funded by Service Public de Wallonie (SPW), Belgium.

References

1. Zontone, P., Affanni, A., Bernardini, R., Del Linz, L., Piras, A., Rinaldo, R.: Supervised learning techniques for stress detection in car drivers. Adv. Sci. Technol. Eng. Syst. J. **5**(6), 22–29 (2020)
2. Healey, J.A., Picard, R.W.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans. Intell. Transp. Syst. **6**(2), 156–166 (2005)
3. Meteier, Q., et al.: Classification of drivers' workload using physiological signals in conditional automation. Front. Psychol. **12**, 596038 (2021)
4. Chen, L.-L., Zhao, Y., Ye, P.-F., Zhang, J., Zou, J.-Z.: Detecting driving stress in physiological signals based on multimodal feature analysis and kernel classifiers. Expert Syst. Appl. **85**, 279–291 (2017)
5. Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., Koltun, V.: CARLA: an open urban driving simulator. In: 2017 Conference on Robot Learning (CoRL), pp. 1–16 (2017)
6. NeuroKit2 Homepage. https://neurokit2.readthedocs.io/
7. Knight, S., Heinrich, A.: Different measures of auditory and visual stroop interference and their relationship to speech intelligibility in noise. Front. Psychol. **8**(230) 2017
8. Pungu Mwange, M.-A., Rogister, F., Rukonić, L.: Measuring driving simulator adaptation using EDA. In: Proceedings of the 13th International Conference on Applied Human Factors and Ergonomics (AHFE) (2022)