

Measuring User Responses to Driving Simulators: A Galvanic Skin Response Based Study

Atiqul Islam

*Research School of Computer Science
Australian National University
Canberra, Australia
atiqul.islam@anu.edu.au*

Jinshuai Ma

*Research School of Computer Science
Australian National University
Canberra, Australia
jinshuai.ma@anu.edu.au*

Tom Gedeon

*Research School of Computer Science
Australian National University
Canberra, Australia
tom@cs.anu.edu.au*

Md Zakir Hossain

*Research School of Computer Science
Australian National University
Canberra, Australia
zakir.hossain@anu.edu.au*

Ying-Hsang Liu

*Research School of Computer Science
Australian National University
Canberra, Australia
ying-hsang.liu@anu.edu.au*

Abstract—The use of simulator technology has become popular in providing training, investigating driving activity and performing research as it is a suitable alternative to actual field study. The transferability of the achieved result from driving simulators to the real world is a critical issue considering later real-world risks, and important to the ethics of experiments. Moreover, researchers have to trade-off between simulator sophistication and the cost it incurs to achieve a given level of realism. This study will be the first step towards investigating the plausibility of different driving simulator configurations of varying verisimilitude, from drivers' galvanic skin response (GSR) signals. GSR is the widely used indicator of behavioural response.

By analyzing GSR signals in a simulation environment, our results are aimed to support or contradict the use of simple low-level driving simulators. We investigate GSR signals of 23 participants doing virtual driving tasks in 5 different configurations of simulation environments. A number of features are extracted from the GSR signals after data preprocessing. With a simple neural network classifier, the prediction accuracy of different simulator configurations reaches up to 90% during driving. Our results suggest that participants are more engaged when realistic controls are used in normal driving, and are less affected by visible context during driving in emergency situations. The implications for future research are that for emergency situations realistic controls are important and research can be conducted with simple simulators in lab settings, whereas for normal driving the research should be conducted with full context in a real driving setting.

Index Terms—Driving simulator, verisimilitude, physiological signal, galvanic skin response, classification

I. INTRODUCTION

Driving simulators have been used for research and training for the past few decades [1], and the complexity of driving simulators have been increasing day by day. Driving simulators differ in fidelity, quality and cost. Some driving simulators in research institutes are built upon real cars and so could have very high levels of verisimilitude, for example the advanced driving simulator used in [2], while others may only consist of simplified components of a car and displays.

Driving simulators have been widely used for investigating drivers' psychological response. One very common type of research is the investigation of human factors [3]. Driver's emotion recognition while driving is important for designing in-vehicle instrument to ensure a safe driving [4]. Usually this kind of research aims at observing and analyzing a driver's behaviour and mental states from the perspectives of psychology, biomedical science and neuroscience. For example authors in [5] investigated the recognition of emotional state of driver through speech interaction between driver and car. Authors in [6] investigated drivers' psychology under the influence of different ice-snow road condition based on a simulator study. Results showed that under such condition drivers' physiological load increases and creates an unfavourable driving situation. Another driving simulator based study showed drivers' risk perception while following another vehicle [7]. This kind of research uses simple driving simulators as the focus is mainly on drivers' physiological signals, and implicitly considering that the details of the car has no significant influence on the results. However, the experiment environment is still important as human behaviour and mental states could be very sensitive to the context and any differences of environment [8], [9].

This leads to a concern whether simple driving simulators can really simulate real driving well and provide drivers with a comparable environment, in which the drivers can have similar physiological behavioural responses as they would in the real world, while driving a real car. Even if this is considered to be true, the question would still remain of the relative importance between the realism of the controls or the amount of visible context in such simple simulator setups.

As human factors research tools, driving simulators must possess some suitable level of verisimilitude. Much research has been completed to validate the performance of high-level driving simulators. Some result shows that for a high-level driving simulator which is built upon real cars, the

human's physiological signals are very close to real driving environments [3], [10], [11]. However, only a few studies have focused on the real world verisimilitude of simple driving simulators, yet many studies like [12], [13] are based on experiments that use simple driving simulators. Thus, it is important to investigate the user response to simple driving simulators in human physiological studies.

In our study, we use five different configurations of simple low-level driving simulators using two different types of controls (keyboard and a force feedback driving set) and three types of display settings (single monitor, triple monitor for large context display and a virtual reality headset). Users' galvanic skin response (GSR) [14] data have been taken during driving in these five different configurations as the GSR signal is the indication of cognitive arousal [15]. We suggest that the verisimilitude of a driving simulator can be measured by the relative reactions of participants in the different settings. This allows us to correlate the effects produced from the users' response to a driving simulator as a result of variations of the driving environment (traffic conditions, number of vehicle, pedestrians and other objects) to the different settings.

The remainder of the paper is organized as follows: Section II gives an account of previous work. Section III discusses the detail of our experimental procedure including the apparatus, methods and analysis of this study. Our new and exciting results are described in Section IV. Finally, Section V provides our conclusions.

II. PREVIOUS WORK

Driving simulators are currently undergoing enormous change with increasing demands for advanced and sophisticated simulators, and the applications range from entertainment to research and advanced training. Based on fidelity, driving simulators can be classified as low-level and high-level. A low-level driving simulator can be illustrated by two parts of this equipment, the driving tool and the visual display. In practical application, low level driving simulators are widely use, especially by the human factors research community as their major focus is human cognitive states. There are many real-world research examples using simple driving simulators to investigate drivers' physiological characteristics [6], [7], [16]–[18]. Most of the studies of this kind assume that the driver's physiological signals in the simulation environment are similar to what they would be in the real world driving environment.

Several studies have focused on high-fidelity simulator validation and comparison with the real world demonstrating similar order results to the simulator results [19], [20]. On the other hand several other studies have focused on validating low-level simulators with relatively higher level ones [21], [22]. Low-level and higher-level simulators are again compared to determine drivers' distraction when using in-vehicle interactive systems [23], [24]. Results suggest that the effects are in similar ranges for both types of simulator. This finding supports the use of low-level simulator in human factor research.

Literature shows the effect of the control has a vital impact on the driving simulator. Studies have shown the influence of steering complexity on simulator realism by subjective assessment. The authors of the paper [25] found strong relation between the steering controller feel and simulator verisimilitude. This result can be useful in selecting steering complexity for better steering feel in simulators. Two different types of steering wheels are compared on a fixed base driving simulator in [26]. The importance of steering haptic feedback on driving performance is investigated by a tractor-driving simulator in [27]. The results show that with no steering force feedback, subjects' controllability decreases and it increases with the increase of the steering torque up to a certain optimal level. A factor that has not been reported in low-level simulators is the effect of controls and display in the same simulator.

Field of view in driving simulators also effects simulator fidelity [28]. Driving simulators with high-level display systems deliver some benefits [29]. This is the reason that many simulation studies have used wider field of view display. However, simulator discomfort remains a concern with high fidelity displays [30]. Moreover, different level of field of view with respect to driving simulator realism was not studied adequately. As the visible context is believed to be an important factor for immersion many studies in virtual reality field investigated the connection between immersion, coherence and physiological signals including GSR [31], [32]. The author in [33] investigated immersive and non-immersive virtual reality (VR) environment using head mounted display (HMD) and laptop computer display respectively but not for driving simulation. They found no significant effect of virtual environment platform on other variables such as sense of presence, task performance and immersion. However, their findings showed that participants' experienced higher simulator sickness with HMD. This study could be an important evidence for supporting or not supporting particular user response to driving simulator using VR headset as a display.

The combined effect of control and display was not adequately studied in the literature. To compare among different controls and visual contexts in different driving situations allows us to determine their relative importance and if there are compensatory influences (such as large field of view compensating for low quality controls, and so on). Galvanic skin responses of a human body serves as a basis for many cognitive, human factor and emotion research [34]–[37]. GSR signal is also used for drivers' cognitive workload detection in driving simulators [38], [39]. Several on-road studies also used GSR signals for driving performance and cognitive workload determination [40], [41]. Many studies compared GSR signals during simulator driving and real-life on-road driving [42]. Based on the literature it is evident that GSR signals could also be useful in determining the level of engagement, stress and cognitive workload in driving simulators. However, using GSR signals for measuring the user response to different driving simulators is not adequately covered in the literature. This motivates us to investigate GSR signal in different simulator environment. Moreover, which part of the simulator setup has

the most effect on different driving scenarios should also be investigated. This paper will focus on investigating driving simulator verisimilitude based on participants' GSR signals for driving in different simulator configurations, driving events and driving scenarios.

III. EXPERIMENT

This study was designed to investigate user responses on different driving simulator settings, driving events and situations based on user GSR signal. An experiment was conducted on simple driving simulator with different configurations.

A. Participants

Twenty-three university students (9 male, 14 female) participated in the experiment with an average age distribution of 20.9 ± 2.3 (mean \pm standard deviation) years. Participants gained course credits for participating in the experiment, but their participation was voluntary - there were a large range of other experiments they could have chosen. All the participants had normal or corrected to normal vision. We only took voluntary participation in the experiment, as the voluntary participants provide highly reliable outcomes compared to paid ones, when they complete experiment tasks [43]. All participants completed the experiment. Before participation, they signed an informed consent form. The experiment was approved by the University's Human Research Ethics Committee.

B. Apparatus

1) *Driving simulator*: The simulator hardware consisted of a computer, displays and controllers. The computer was a Windows 10 based machine with Intel Core i7 3770 CPU, Nvidia GTX 1060 6 GB graphic card and 16 GB RAM. A commercial driving software City Car Driving (CCD) was used, which was designed for new drivers or learners to practice driving skills [44]. CCD provides a number of flexible options to simulate various driving environment like different gear styles, traffic density etc. We set the gear style to automatic and thus only two gears were used (forward and backward) as our focus was GSR signals in different settings and this removes the cognitive load of changing gears often as would be the case with a manual gear setting. To trigger emergency situations, the traffic level was set to 70%; and the emergency level was set to "very often" for all driving trials. Even on this setting, emergency situations were a relatively low proportion of the driving time. Two driving situation classes were labelled, which were normal driving situations and emergency driving situations.

The normal driving situation was when participants drove the virtual vehicle forward normally, stopped for traffic and so on. Emergency situations were when an accident happened which included hitting or almost hitting another car, a pedestrian or an object. According to the feedback and results from pilot experiments, our configurations were optimal so that it balances normal situations and emergencies, and the participants were able to experience these stimulations without

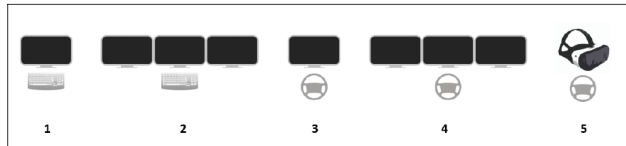


Fig. 1: Configurations of the simulator

TABLE I: Configurations of the simulator

Task No.	Configuration Name	Controller	Display
Task 1	KS	Keyboard	Single Monitor
Task 2	KT	Keyboard	Triple Monitor
Task 3	DS	Driving Set	Single Monitor
Task 4	DT	Driving Set	Triple Monitor
Task 5	DH	Driving Set	Headset (Virtual Reality)

getting either bored or annoyed during the 45 minutes long experiment.

We used low-level driving simulator in our study as our focus was to support or contradict the use of simple low-level driving simulators in real world research. Then we tried to increase the amount of realism by adding more display units and realistic controls but still keeping the simulator in low-level form. Usually the low-level simple driving simulator consists of the display and controller. In this experiment, to create several different configurations of simple driving simulator with different level of realism, the displays were divided into three groups. The first was a single 24 inch computer monitor with standard full high definition (FHD) resolution. The second was a triple monitor set, each with the same specifications. The third was a Fove virtual reality (VR) headset with a standard quad high definition (QHD) resolution.

The controllers were divided into two groups. The first was a standard computer keyboard. The second was a Logitech driving set consisting of a driving steering wheel, a pedal set and a shifter. Using different combination of displays and controllers, a total of 5 configurations had been designed for the simulator, as shown in Figure 1 and Table I. We interchangeably used "tasks" and "configurations" to indicate driving in different simulator configuration setups.

2) *Sensors*: Participants' GSR signals were collected by the E4 wristband device produced by Empatica (<https://www.empatica.com/>) which is capable of collecting real-time GSR signal at a sampling rate of 4 Hz, along with various other data. This device is getting more and more popularity these days among human factor researcher. Many studies have used Empatica device with well accepted accuracies [45]–[47]. We used the Fove (<https://www.getfove.com/>) virtual reality (VR) headset with integrated eye tracker, which is designed for gaming and eye feature research in VR environments.

C. Experiment Procedure

Participants were briefed about the experiment and asked to read the information sheet, after that they completed the



Fig. 2: Overview of the simulator setup

ANU approved consent form. After the initialization and the calibration of sensors, the participants were asked to practice for 3 minutes to get used to and comfortable with the experiment environment and devices. Participant could move the non-rotating chair forward or backward to adjust their best suited position. They were asked to minimize any unnecessary movement except the natural movement for completing the task. So that the movement related artefact could be minimized. After the practice session on the simulator, the participants performed the 5 driving tasks using different simulator configurations.

At the beginning of each task, the simulator setup was switched to the appropriate one, and using the appropriate devices. To make it convenient to switch between different configurations, the monitors were put together, and the driving set and keyboard were put on the same table as shown in Figure 2. Between consecutive tasks in the experiment, there is a 2 minutes break in which the participants relax. At the end, each participant was asked to fill in a questionnaire, which included basic demographic questions and some questions about participants' driving experience and feedback about the simulator setups.

Participants' familiarity with the devices may potentially affect their GSR signals. For example, a participant may perform better in the last task than the first one because the user gradually gets used to the devices and becomes more skilful in operating them. To minimise the potential learning effect of the assigned tasks, the five tasks in each experiment were conducted in an order balanced fashion. We followed a Latin square method for this [48], [49].

D. Data Analysis

The analysis of the sensor data was performed in three steps: signal pre-processing, feature extraction, and analysis. Pre-processing involved signal labelling, signal filtering to remove noise artefacts and normalization to reduce subject dependency. Feature extraction involved segmentation and

TABLE II: Driving events and situations in each experiment session

Events	Situations
Hit a pedestrian	Emergency
Almost hit a pedestrian	
Hit an object	
Almosyt hit an object	
Hit a car	
Almost hit a car	Normal
Normal driving	
Stopping	

feature calculation. The analysis used in this study involved simple statistical methods and neural network classification.

A manual labelling program was developed and implemented to label eight different driving events during each experiment session. The events are also divided in two situation categories as shown in the Table II. Each label will be logged with a unique millisecond timestamp. The program will also record the beginning and end of each driving task. A 20 point median filter was used for the signals, which removed outliers and smoothed the signal but retained its original shape [50].

GSR signals have variable ranges of baseline values depending on different subjects, which means the mean value of a signal may have differences between different individuals. So, normalization was done on the data over the whole period of a particular participant to scale different participants' signals into the same range considering each participant separately [51]. We used maximum value normalization. In this normalization, each value of a participant's GSR signal for one complete experiment is divided by the maximum value of that particular participant's GSR signal [52], [53]. Thus, signals for each participant varied between 0 and 1 overall, and not for each simulator configuration.

Afterward, the data was first segmented based on each task which was actually the simulator configuration (see Figure 1 and Table I), for example, configuration DT is one segment and similarly other configurations. In each task, the data was segmented again based on the events as shown in Table II. During each driving tasks, a number of driving events are labelled. These labels are used for segmentation. Each label contains a unique timestamp and 5 seconds of data segment around the timestamp is extracted which spans from one second before the unique timestamp of the label to four seconds after the timestamp. A number of temporal and spectral features were then calculated from each segment. In time domain, maximum, minimum, mean, median value, standard deviation, variance, number of zero crossings, root mean square were used as features. Fast Fourier Transform (FFT) [54] was applied on the original data to convert it to the frequency domain signal. The skewness and kurtosis are applied on frequency domain signals. We refer the reader to [55] for the equations of the features.

To investigate the differences in users' GSR signals to different driving environments in our study, an analysis of

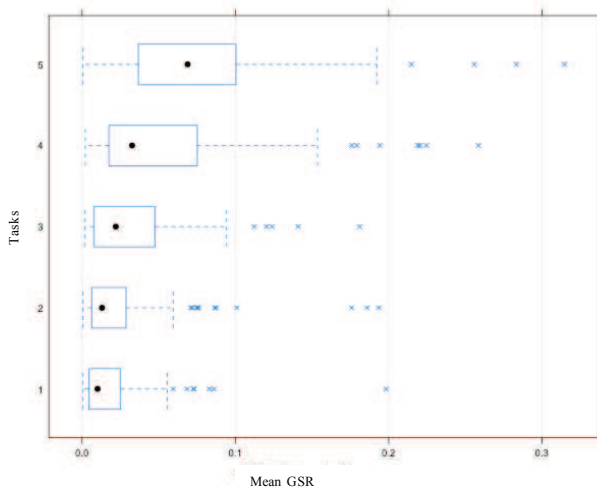


Fig. 3: Relationship between the mean GSR and task (Task 1 = single monitor + keyboard (KS); Task 2 = triple monitor + keyboard (KT); Task 3 = single monitor + driving set (DS); Task 4 = triple monitor + driving set (DT); and Task 5 = driving set + headset virtual reality (DH)).

variance (ANOVA) and TukeyHSD ad-hoc comparisons were applied on the mean absolute value of GSR signals under different simulator configurations [56]. The GSR signals under normal driving and emergency driving situations were analyzed independently. However, the method can only detect significant differences without deciding where and what the differences are [57]. To further investigate the differences, a neural network was applied on the selected features to perform classification on the 5 simulator configurations respectively. Artificial neural networks have been widely applied to classify physiological data for emotion classification such as fake and genuine smile detection [50] and has been shown to have good performance. Classification result shows the accuracy in classifying different configurations of the simulator based on GSR signals. The network architecture included 10 hidden neurons, tan-sigmoid activation function and mean squared normalized error performance function(mse). The input were the selected GSR features and there were two output classes, the normal situation and emergency situation. A leave-one-out approach was applied to validate the performance, which used each subject's data as testing set while others as training set. Therefore the validation ran a total of 23 loops, and the result was the average of each loop. The leave-one-out validation process was repeated for 10 times and we average the results to get the final result.

IV. RESULTS AND DISCUSSION

Firstly, significance differences were tested between mean GSR and a condition (situation, event, or task) using an analysis of variance (ANOVA) technique. From the analysis, we did not find any significant relationship between the mean GSR and the situation ($F(1,591)= 1.49, p = .22, p > .05$), as

TABLE III: Relationship between user characteristics and mean GSR at 95% confidence interval

User Characteristics	Cut Point (Mean)	Odds Ratio	Log Odds	Std Error	t-Value	Significant?
Gender	0.04	1.71	0.54	0.18	2.92	Yes
Wear glasses	0.04	0.59	-0.53	0.20	-2.73	Yes
Driving license	0.04	0.45	-0.80	0.29	-2.72	Yes
Driving simulator	0.04	0.48	-0.74	0.31	-2.39	Yes
Devices affecting performance	0.04	0.69	-0.37	0.18	-2.07	Yes
See through VR headset clearly	0.04	1.66	0.51	0.18	2.88	Yes

well as between the mean GSR and the event ($F(7,585) = 1.24, p = .28, p > .05$). However, extremely significant relationships were observed between the mean GSR and the task (i.e. each configuration) ($F(4,588) = 63.85, p = <2e-16, p < .001$). The values of mean GSR (considering each task separately) are shown in Figure 3.

Secondly, TukeyHSD was applied to measure the significant differences between task pairs. The results show that there are significant differences between the task pairs of 3-1, 4-1, 5-1, 3-2, 4-2, 5-2, 4-3, 5-3 and 5-4. It illustrates that although there are significant differences between task 1 and other tasks (tasks 3, 4 and 5), as well as between task 2 and other tasks (tasks 3, 4 and 5), but there is no significant difference between task 1 (single monitor + keyboard) and task 2 (triple monitor + keyboard). It is not surprising as people's focus narrows during sudden stress. So it is plausible that it matters little if the participant is looking at a single screen or a triple screen display. The result is also evident from a study where significant differences were not found for measuring the sense of presence considering display context [58].

Finally, we find significant differences in the mean GSR for all the three tasks with driving set (4-3, 5-3 and 5-4), and in the mean GSR between the keyboard and the driving set settings. Within the driving set setting, there exist significant differences in the mean GSR for different configurations, i.e., single monitor, triple monitor and VR headset. Overall, the results show that there are significant differences in GSR considering most of the tasks in the study.

Afterward, the association between user characteristics and the mean GSR was evaluated using an odds ratio analysis [59]. The results are summarised in Table III. According to the analysis, it was observed that female participants are more likely to have higher mean GSR than male participants. People who wear glasses are more likely to have lower mean GSR than people who do not wear glasses. People who feel that devices affect driving performance are more likely to have lower mean GSR. People who can see everything clearly through the VR headset are more likely to have higher mean GSR. People who have experience using a driving simulator are more likely to have lower mean GSR. Overall, we found significant differences considering mean GSR and six user characteristics separately as shown in Table III. It shows that GSR is a good indicator to differentiate between different driv-

TABLE IV: Classification result based on simulator configurations

Name of the configurations	Classification Accuracy
Keyboard-Single Monitor (KS)	0.636
Keyboard-Triple Monitor (KT)	0.809
Driving Set-Single Monitor (DS)	0.873
Driving Set-Triple Monitor (DT)	0.900
Driving Set-Headset VR (DH)	0.773

ing stresses levels when considering user characteristics such as demographics, previous experiences and user perceptions.

In addition, classification results for all 5 different configurations using all features extracted from the signals are shown in Table IV. The value in the table is the prediction accuracy between normal and emergency driving for each class. Here classification accuracy is the indication of strong impact by the configuration setups on the GSR readings. As verisimilitude increases participants' engagement with the simulator will increase. So, the classifier indicates the degree of engagement in different setups. The table shows that the prediction accuracies of KS and KT are lower, being 64% and 81% respectively where the controller for the simulator is only a keyboard. For a driving simulator this is very basic and there is little realism in a keyboard to control a car. When we replace the keyboard by a driving set with a force-feedback steering wheel, gear shifter, brake and accelerator the accuracy increases for DS and DT to 87% and 90% respectively. This is also supported in [60] where a driving video game was played with different controllers as an independent variable. Their results supported that a driving wheel controller being perceived as more natural than keyboard, gamepad, or joystick controllers.

We should mention that the VR headset shows a prediction accuracy of 77% (see Table IV). We suggest that the VR headset is not 'real' compared to driving simulator in all ways, as the participants could not see their own hands or their own controlling device in the simulator, which may have had effects on their perception of verisimilitude and hence on the GSR signals of participants. Thus, we propose a controller verisimilitude model, excluding DH, as shown in Figure 4. It represents participants' spatial presence considering the position of real-time simulator configurations. Please note that the figure is not drawn to scale.

Our two dimensional model suggests that there is possibly some combined effect of the control device and the degree of view context, and also shows that the control device is substantially more important as the percent difference is very small (87% to 90%) in the two settings using the steering set. Our results suggests that when a virtual reality headset is used, the verisimilitude of the simulator is lacking when compared to the monitor display, and showing 77% which is a little above the keyboard-single monitor configuration but below all other configurations. Although VR is supposed to

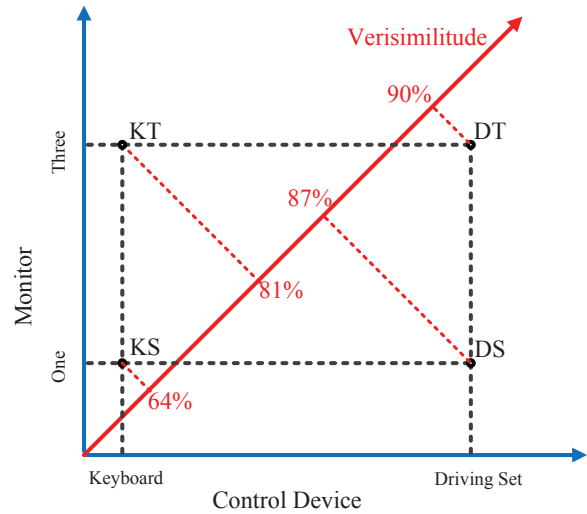


Fig. 4: A two dimensional controller verisimilitude model (positions are not drawn to scale; KS = Keyboard-Single Monitor; KT = Keyboard-Triple Monitor; DS = Driving Set-Single Monitor; DT = Driving Set-Triple Monitor).

be the most immersive environment in gaming contexts, in a driving simulator it lacks some verisimilitude due to the participants' feeling of discomfort and lack of visible arms. The literature suggests that the negative impact of discomfort reduces the immersion effect of VR [61]. Other researcher also supports this claim that head mounted display elicit more simulation sickness than standard desktop displays [33].

V. CONCLUSIONS

The current study sought to investigate the relationship between drivers' GSR signals and response to different verisimilitude levels of simple driving simulators.

Participants' GSR signals are captured at 5 different levels of verisimilitude of simulator. Simple driving simulation environments are mainly used by researchers, particularly psychology and human-computer interaction (HCI) researchers to study drivers' behaviour and mental states in different conditions. However, whether the GSR signal aroused by simple simulation environments is close to real driving environment or not remains a question. Also research in dangerous and emergency driving situation always remains out of scope of real experiment as the research ethics never permit this kind of experiment, driving simulators play a significant role in research based on driving in emergencies.

Our experiment with 23 participants undertaking a series of driving tasks using simulator setups with different level of verisimilitude attempted to answer this question. Participants' GSR data was collected and analyzed. Both statistical method and neural network classification approaches were used. The significance test analyzed the relation between the mean value of each GSR signal and the verisimilitude level of simulator.

The classification approach focused on extracted features and analyzes all participants together as whole.

Driving in emergency situations is explicitly believed to be more stressful than driving in normal situations. However, both statistical and classification results show that in emergency situations, the verisimilitude of the simulator has less influence on GSR signals, as compared to normal driving situations. We explain this as the stress in emergency situations is much stronger than the effect on GSR readings due to change of configuration setups i.e. verisimilitude. This implies that simple driving simulators can be used for human factor research involving emergency situations where real life experiments are ethically not permitted. Also, we found that the differences of controllers affect the GSR signal more strongly than the visible context. The VR headset setting also has a significant influence, however, it is not easy to place with the other four configurations, in our future work we will investigate further. Also we will investigate some other physiological signals such as blood volume pulse, heart rate, pupillary measures, and we will also increase the number of simulator configurations, by for example including the body of a car and use a wider range of labels for different kinds of emergency situations. In our future work, we will increase the size of the data set, although twenty or more participants are considered as enough in research based on physiological signals [62].

ACKNOWLEDGMENT

Authors want to thank all the participants who voluntarily participated in this experiment.

REFERENCES

- [1] E. Blana, "A survey of driving research simulators around the world." 1996.
- [2] G. S. Larue, A. Rakotonirainy, and N. L. Haworth, "A simulator evaluation of effects of assistive technologies on driver cognitive load at railway-level crossings," *Journal of Transportation Safety & Security*, vol. 8, no. suppl, pp. 56–69, 2016.
- [3] B. Philips and T. Morton, "Making driving simulators more useful for behavioral research-simulator characteristics comparison and model-based transformation: Summary report," Tech. Rep., 2015.
- [4] M. Grimm, K. Kroschel, H. Harris, C. Nass, B. Schuller, G. Rigoll, and T. Moosmayr, "On the necessity and feasibility of detecting a drivers emotional state while driving," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2007, pp. 126–138.
- [5] C. M. Jones and M. Jonsson, "Detecting emotions in conversations between driver and in-car information systems," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2005, pp. 780–787.
- [6] L. W. Hu and Y. L. Pei, "Analysis of driver's physiological characteristics under ice-snow conditions base on driving simulator," in *Applied Mechanics and Materials*, vol. 109. Trans Tech Publ, 2012, pp. 547–550.
- [7] H. Nakamura, S. Yamabe, K. Nakano, D. Yamaguchi, and Y. Suda, "Driver risk perception and physiological state during car-following experiments using a driving simulator," *International Journal of Intelligent Transportation Systems Research*, vol. 8, no. 3, pp. 140–150, 2010.
- [8] J.-W. Lin, H. B.-L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness, "Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment," in *Virtual Reality, 2002. Proceedings. IEEE*. IEEE, 2002, pp. 164–171.
- [9] R. P. McMahan, D. A. Bowman, D. J. Zielinski, and R. B. Brady, "Evaluating display fidelity and interaction fidelity in a virtual reality game," *IEEE Transactions on Visualization & Computer Graphics*, no. 4, pp. 626–633, 2012.
- [10] J. Li, X. Zhao, S. Xu, J. Ma, and J. Rong, "The study of driving simulator validation for physiological signal measures," *Procedia-Social and Behavioral Sciences*, vol. 96, pp. 2572–2583, 2013.
- [11] M. Gemou, "Transferability of driver speed and lateral deviation measurable performance from semi-dynamic driving simulator to real traffic conditions," *European Transport Research Review*, vol. 5, no. 4, pp. 217–233, 2013.
- [12] A. Anund, C. Fors, and C. Ahlstrom, "The severity of driver fatigue in terms of line crossing: a pilot study comparing day-and night time driving in simulator," *European transport research review*, vol. 9, no. 2, p. 31, 2017.
- [13] P. Choudhary and N. R. Velaga, "Analysis of vehicle-based lateral performance measures during distracted driving due to phone use," *Transportation research part F: traffic psychology and behaviour*, vol. 44, pp. 120–133, 2017.
- [14] W. Boucsein, *Electrodermal activity*. Springer Science & Business Media, 2012.
- [15] F. G. Paas, J. J. Van Merriënboer, and J. J. Adam, "Measurement of cognitive load in instructional research," *Perceptual and motor skills*, vol. 79, no. 1, pp. 419–430, 1994.
- [16] K. A. Brookhuis and D. de Waard, "Monitoring drivers' mental workload in driving simulators using physiological measures," *Accident Analysis & Prevention*, vol. 42, no. 3, pp. 898–903, 2010.
- [17] E. Ismail, A. S. Ghazali, and S. N. Sidek, "Physiological signalbased engagement level analysis under fuzzy framework," in *Applied Mechanics and Materials*, vol. 373. Trans Tech Publ, 2013, pp. 1768–1775.
- [18] C. Zhao, M. Zhao, J. Liu, and C. Zheng, "Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator," *Accident Analysis & Prevention*, vol. 45, pp. 83–90, 2012.
- [19] X. Yan, M. Abdel-Aty, E. Radwan, X. Wang, and P. Chilakapati, "Validating a driving simulator using surrogate safety measures," *Accident Analysis & Prevention*, vol. 40, no. 1, pp. 274–288, 2008.
- [20] F. Bella, "Driving simulator for speed research on two-lane rural roads," *Accident Analysis & Prevention*, vol. 40, no. 3, pp. 1078–1087, 2008.
- [21] M. Klüver, C. Herrigel, C. Heinrich, H.-P. Schöner, and H. Hecht, "The behavioral validity of dual-task driving performance in fixed and moving base driving simulators," *Transportation research part F: traffic psychology and behaviour*, vol. 37, pp. 78–96, 2016.
- [22] F. Rosey and J.-M. Auberlet, "Driving simulator configuration impacts drivers behavior and control performance: An example with studies of a rural intersection," *Transportation research part F: traffic psychology and behaviour*, vol. 27, pp. 99–111, 2014.
- [23] G. Burnett, A. Irune, and A. Mowforth, "Driving simulator sickness and validity: how important is it to use real car cabins?" *Advances in transportation studies*, vol. 2007, no. Spec Iss, pp. 33–42, 2007.
- [24] S. L. Jamson and A. H. Jamson, "The validity of a low-cost simulator for the assessment of the effects of in-vehicle information systems," *Safety Science*, vol. 48, no. 10, pp. 1477–1483, 2010.
- [25] B. Shyrokau, J. Loof, O. Stroosma, M. Wang, and R. Happee, "Effect of steering model fidelity on subjective evaluation of truck steering feel," *DSC EUROPE*, 2015.
- [26] R. R. Mourant and P. Sadhu, "Evaluation of force feedback steering in a fixed based driving simulator," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 46, no. 26. SAGE Publications Sage CA: Los Angeles, CA, 2002, pp. 2202–2205.
- [27] T. M. Grant, D. Karimi, and D. D. Mann, "Effect of steering torque feedback on driving performance in a tractor simulator," in *ASABE/CSBE North Central Intersectional Meeting*. American Society of Agricultural and Biological Engineers, 2006, p. 1.
- [28] H. Jamson, "Driving simulation validity: issues of field of view and resolution," in *Proceedings of the driving simulation conference*, 2000, pp. 57–64.
- [29] V. Govil, S. Lovell, P. Suresh, Q. Wu, G. Yang, and R. R. Mourant, "A high-resolution wide-screen display for simulators and virtual reality," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 48, no. 17. SAGE Publications Sage CA: Los Angeles, CA, 2004, pp. 2131–2133.
- [30] Ł. Dziuda, M. P. Biernacki, P. M. Baran, and O. E. Truszczyński, "The effects of simulated fog and motion on simulator sickness in a driving simulator and the duration of after-effects," *Applied ergonomics*, vol. 45, no. 3, pp. 406–412, 2014.
- [31] M. Meehan, B. Insko, M. Whitton, and F. P. Brooks Jr, "Physiological measures of presence in stressful virtual environments," in *Acm transactions on graphics (tog)*, vol. 21, no. 3. ACM, 2002, pp. 645–652.

- [32] R. Skarbez, F. P. Brooks Jr, and M. C. Whitton, "Immersion and coherence in a stressful virtual environment," in *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. ACM, 2018, p. 24.
- [33] S. Khashe, B. Becerik-Gerber, G. Lucas, and J. Gratch, "Persuasive effects of immersion in virtual environments for measuring pro-environmental behaviors," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, vol. 35. IAARC Publications, 2018, pp. 1–7.
- [34] Y. Shi, N. Ruiz, R. Taib, E. Choi, and F. Chen, "Galvanic skin response (gsr) as an index of cognitive load," in *CHI'07 extended abstracts on Human factors in computing systems*. ACM, 2007, pp. 2651–2656.
- [35] Y. Nagai, C. I. Jones, and A. Sen, "Galvanic skin response (gsr)/electrodermal/skin conductance biofeedback on epilepsy: a systematic review and meta-analysis," *Frontiers in neurology*, vol. 10, p. 377, 2019.
- [36] R. Phitayakorn, R. Minehart, M. Pian-Smith, M. Hemingway, and E. Petrusa, "Practicality of using galvanic skin response to measure intraoperative physiologic autonomic activation in operating room team members," *Surgery*, vol. 158, no. 5, pp. 1415–1420, 2015.
- [37] G. I. Christopoulos, M. A. Uy, and W. J. Yap, "The body and the brain: measuring skin conductance responses to understand the emotional experience," *Organizational Research Methods*, vol. 22, no. 1, pp. 394–420, 2019.
- [38] W. Hajek, I. Gaponova, K. Fleischer, and J. Krems, "Workload-adaptive cruise control—a new generation of advanced driver assistance systems," *Transportation research part F: traffic psychology and behaviour*, vol. 20, pp. 108–120, 2013.
- [39] J. Son, B. Reimer, B. Mehler, A. Pohlmeier, K. Godfrey, J. Orszulak, J. Long, M. Kim, Y. Lee, and J. F. Coughlin, "Age and cross-cultural comparison of drivers' cognitive workload and performance in simulated urban driving," *International Journal of Automotive Technology*, vol. 11, no. 4, pp. 533–539, 2010.
- [40] J. Healey, R. W. Picard *et al.*, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.
- [41] G. Jahn, A. Oehme, J. F. Krems, and C. Gelau, "Peripheral detection as a workload measure in driving: Effects of traffic complexity and route guidance system use in a driving study," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, no. 3, pp. 255–275, 2005.
- [42] B. Reimer and B. Mehler, "The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation," *Ergonomics*, vol. 54, no. 10, pp. 932–942, 2011.
- [43] J. Redi and I. Pova, "Crowdsourcing for rating image aesthetic appeal: Better a paid or a volunteer crowd?" in *Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia*. ACM, 2014, pp. 25–30.
- [44] F. Development, "City Car Driving - car driving simulator, pc game," <https://citycardriving.com/>, 2018, accessed: 10 August 2018.
- [45] H. Koskimäki, H. Mönttinen, P. Siirtola, H.-L. Huttunen, R. Halonen, and J. Rönning, "Early detection of migraine attacks based on wearable sensors: experiences of data collection using empatica e4," in *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. ACM, 2017, pp. 506–511.
- [46] C. McCarthy, N. Pradhan, C. Redpath, and A. Adler, "Validation of the empatica e4 wristband," in *2016 IEEE EMBS International Student Conference (ISC)*. IEEE, 2016, pp. 1–4.
- [47] R. A. Sarkis, S. Thome-Souza, M.-Z. Poh, N. Llewellyn, J. Klehm, J. R. Madsen, R. Picard, P. B. Pennell, B. A. Dworetzky, T. Loddenkemper *et al.*, "Autonomic changes following generalized tonic clonic seizures: an analysis of adult and pediatric patients with epilepsy," *Epilepsy research*, vol. 115, pp. 113–118, 2015.
- [48] A. L. Edwards, "Balanced latin-square designs in psychological research," *The American journal of psychology*, vol. 64, no. 4, pp. 598–603, 1951.
- [49] J. T. Richardson, "The use of latin-square designs in educational and psychological research," *Educational Research Review*, vol. 24, pp. 84–97, 2018.
- [50] M. Z. Hossain, T. Gedeon, and R. Sankaranarayana, "Using temporal features of observers' physiological measures to distinguish between genuine and fake smiles," *IEEE Transactions on Affective Computing*, 2018.
- [51] J. T. Cacioppo and L. G. Tassinary, "Inferring psychological significance from physiological signals." *American psychologist*, vol. 45, no. 1, p. 16, 1990.
- [52] D. T. Lykken and P. H. Venables, "Direct measurement of skin conductance: A proposal for standardization," *Psychophysiology*, vol. 8, no. 5, pp. 656–672, 1971.
- [53] M. Z. Hossain, T. Gedeon, and A. Islam, "Understanding two graphical visualizations from observer's pupillary responses and neural network," in *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. ACM, 2018, pp. 215–218.
- [54] W. T. Cochran, J. W. Cooley, D. L. Favon, H. D. Helms, R. A. Kaenel, W. W. Lang, G. C. Maling, D. E. Nelson, C. M. Rader, and P. D. Welch, "What is the fast fourier transform?" *Proceedings of the IEEE*, vol. 55, no. 10, pp. 1664–1674, 1967.
- [55] D. D. Ayata, Y. Yaslan, and M. Kamaşak, "Emotion recognition via galvanic skin response: Comparison of machine learning algorithms and feature extraction methods," *Istanbul University-Journal of Electrical & Electronics Engineering*, vol. 17, no. 1, pp. 3147–3156, 2017.
- [56] P. K. Dunn and G. K. Smyth, *Generalized linear models with examples in R*. Springer, 2018.
- [57] A. Aron and E. N. Aron, *Statistics for psychology*. Prentice-Hall, Inc, 1994.
- [58] J. Kuusisto, A. Ellman, T. Kaapu, and T. Tiainen, "Effect of the immersion level of a virtual loader simulator on the sense of presence," in *ASME 2011 World Conference on Innovative Virtual Reality*. American Society of Mechanical Engineers, 2011, pp. 155–163.
- [59] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical methods for rates and proportions*. John Wiley & Sons, 2013.
- [60] P. Skalski, R. Tamborini, A. Shelton, M. Buncher, and P. Lindmark, "Mapping the road to fun: Natural video game controllers, presence, and game enjoyment," *New Media & Society*, vol. 13, no. 2, pp. 224–242, 2011.
- [61] D. Michael, M. Kleanthous, M. Savva, S. Christodoulou, M. Pampaka, and A. Gregoriades, "Impact of immersion and realism in driving simulator studies," *International Journal of Interdisciplinary Telecommunications and Networking (IJITN)*, vol. 6, no. 1, pp. 10–25, 2014.
- [62] M. Z. Hossain and T. Gedeon, "Observers' physiological measures in response to videos can be used to detect genuine smiles," *International Journal of Human-Computer Studies*, vol. 122, pp. 232–241, 2019.