

Research Article

Comparing Expert Driving Behavior in Real World and Simulator Contexts

**Hiran B. Ekanayake,^{1,2} Per Backlund,³ Tom Ziemke,³ Robert Ramberg,¹
Kamalanath P. Hewagamage,² and Mikael Lebram³**

¹ Stockholm University, Forum 100, 164 40 Kista, Sweden

² University of Colombo, School of Computing, 35 Reid Avenue, 00700 Colombo 7, Western Province, Sri Lanka

³ University of Skövde, P.O. Box 408, 541 28 Skövde, Sweden

Correspondence should be addressed to Hiran B. Ekanayake; hbe@ucsc.lk

Received 21 February 2013; Revised 13 June 2013; Accepted 15 July 2013

Academic Editor: Alexander Pasko

Copyright © 2013 Hiran B. Ekanayake et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Computer games are increasingly used for purposes beyond mere entertainment, and current hi-tech simulators can provide quite, naturalistic contexts for purposes such as traffic education. One of the critical concerns in this area is the validity or transferability of acquired skills from a simulator to the real world context. In this paper, we present our work in which we compared driving in the real world with that in the simulator at two levels, that is, by using performance measures alone, and by combining psychophysiological measures with performance measures. For our study, we gathered data using questionnaires as well as by logging vehicle dynamics, environmental conditions, video data, and users' psychophysiological measurements. For the analysis, we used several novel approaches such as scatter plots to visualize driving tasks of different contexts and to obtain vigilance estimators from electroencephalographic (EEG) data in order to obtain important results about the differences between the driving in the two contexts. Our belief is that both experimental procedures and findings of our experiment are very important to the field of serious games concerning how to evaluate the fitness of driving simulators and measure driving performance.

1. Introduction

There is a growing interest to use simulators for educational and training purposes by using traditional entertainment oriented and personal computers based gaming platforms, which are commonly referred to as serious games [1–3]. For instance, according to SWOV [4], about 150 driving simulators were used for basic driver training in 2010 in the Netherlands. Although driving simulators bring many advantages to driver training, such as safe practice environment and unlimited repetition, there is a question of validity, that is, whether the competence or performance obtained in the simulator is valid in real world driving. To our knowledge, little research has focused on this question (e.g., [5]) because of reasons such as the risk of testing the skills in the real world, higher costs and efforts required in such research, and methodological weaknesses. Addressing this problem, this research has evaluated the equivalence between driving in

the real world and driving in a simulator at two levels of enquiry: by using performance measures alone and by combining psychophysiological measures with performance measures.

For our investigation, we involved experienced drivers and collected data about both real world driving and driving in a mid-range driving simulator. The data were gathered in various forms, that is, quantitative data related to vehicle dynamics (e.g., steering angle), environment conditions (e.g., vehicle speed), and driver's psychophysiological signals (i.e., electroencephalographic and heart rate), questionnaire data, and video data. In the first level of analysis, we compared tasks of real world driving and driving in the simulator using simple graphs as well as using scatter plots, and we found interesting results such as drivers' perceptions about driving vary greatly in the two driving contexts. For analyzing electroencephalographic (EEG) data, we proposed an improved technique to overcome the limitations and challenges, such as

artifacts due to movements of subjects, less accurate equipment, and the fact that there are no psychological indices to directly associate the EEG features with. The purpose of the second level of analysis was to capture hidden physiological influences on drivers' performance in the two driving contexts. As a result, we were able to confirm the findings of the previous level of analysis and to infer further findings to compare driving in the two driving contexts. Although our approach cannot handle issues of transfer of learning in terms of describing implications for traffic education based on comparisons between simulator and real world driving, we are convinced that our research provides important findings for simulator based traffic education concerning how to evaluate fitness of driving simulators and measure driving performance.

The paper is organized as follows. Section 2 discusses the use, such as benefits and requirements, of driving simulators for traffic education, factors that might be affecting the perceived realism of drivers and their performance in such learning environments, and the three types of measures—performance, physiological, and subjective—that can be considered to compare the difference between driving in the real world and driving in a simulator. Section 3 of the paper presents the methodology which includes the justification of our approach, the experimental setup, and the procedure of analysis of performance measures and psychophysiological measures. The results are presented in Section 4. In the discussion section (Section 5), we discuss the findings of the experiment in greater detail. Finally, in Section 6, we conclude by presenting a summary of the findings, limitations of our approach, and suggestions for future work.

2. Related Work

2.1. Driving Simulators for Traffic Education. Driving simulators offer many advantages to traffic education. According to Fuller (as interpreted in [4]), they offer faster exposition to a wide variety of traffic situations, improved possibilities for feedback from different perspectives, unlimited repetition of educational moments, computerized and objective assessments, demonstration of maneuvers, and safe practice environment. They also allow factors closely related to self-efficacy to be adjusted or altered which have a direct effect on the perception of task difficulty, motivation, and locus of control [6], as well as allowing researchers to analyze risky scenarios without endangering a participant [7]. A previous study showed that a game-based simulation can be used to improve traffic safety variables such as speed, use of turn signal and rear-view mirrors, headway distance, and lane change behavior [1]. However, this study has not validated the effects of such learning in real world driving.

Apart from the technical quality, the other important requirements of a simulator for training purposes are the quality of the simulator's lessons, appropriateness of instruction and feedback, and adaptability of simulator lessons to the pace and learning style of the individual learner [4]. Although all these requirements are met, simulation is still an imitation of reality that is far from being perfect. This specific issue links to situated and distributed cognition which identifies

the importance of conducting learning in a meaningful and supportive context, and it identifies problems of transfer if the learning environment deviates considerably from reality [4, 6]. However, attributed to self-efficacy theory, [6] identifies that:

if self-efficacy for driving a car in real life is promoted by driving in a simulator, by making the driver more attentive, judicious, and so forth, as reflected in an actual improvement of performance, then there is learning above the limitations of the simulator.

Moreover, recalling that all simulators are models of what they are simulating, Gee [8] argues that

Models and modeling are important to learning because, although people learn from their interpreted experience, models and modeling allow specific aspects of experience to be interrogated and used for problem solving in ways that lead from concreteness to abstractness.

2.2. Vigilance, the Nature of Task, and Driving Performance.

The previous topic has identified driving simulators as more advantageous learning environments for traffic education, but expecting the learning drivers' active participation during the learning cycle. However, it is equally important to have a look at factors that might be affecting the perceived realism of drivers and their performance in such learning environments, such as the vigilance and how the nature of the task might be affecting them.

Vigilance, also called sustained attention, refers to the ability of organisms to maintain their focus of attention and to remain alert to stimuli over prolonged periods of time [9]. However, vigilance tends to decline, a phenomenon called vigilance decrement, resulting in substantial failures in human performance. For instance, road accidents are often caused by failures of vigilance in drivers [10].

Traditionally, vigilance decrement has been conceived as a decline in arousal as a result of low cognitive demands. A theoretical position that supports this view is called the arousal theory which suggests an "inverted-U" shaped relationship between arousal and task performance [11, 12], that is, task performance is poor when arousal is either too weak or too strong. However, this theory has failed to explain high-stress levels associated with vigilance and underestimated the nature of vigilance task. On the contrary, more recent studies indicate that an individual's vigilance depends on mental resources that can be allocated to a task [9, 13]. Since we are investigating drivers' performance in two different driving contexts, it is worth to elaborate how vigilance is associated with the driving environment and performance of drivers.

Studies investigating the vigilance of drivers report that driving under decreased levels of vigilance will cause longer reaction time, attention decline, and deficits in information processing and will ultimately increase the risks of accidents [10, 11, 14]. As discussed by Thiffault and Bergeron [11], there are two broad conceptions of vigilance: one is associated with

physiological processes which have influence on alertness and wakefulness, and the other is associated with information processing and sustained attention. The same can be understood from the multimodal nature of emotion [15, 16]. Factors influencing the physiological states that underlie vigilance and alertness can be categorized into endogenous and exogenous [11, 17]. Endogenous factors, such as time of day, duration of task, and sleep-related problems, are associated with long-term fluctuations of alertness and affect the basic preparation state of the individual. However, exogenous factors are determined by the individual's interactions with the road environment such as its monotony and low traffic density, and they have an impact on the driving performance by affecting alertness, information processing, and arousal. Since our research mainly focuses on the exogenous factors, that is, each individual's interaction with the two types of road environments, it is important to elaborate the discussion within that scope, such as relations between road scene, speed, vigilance, and driving performance.

The role of speed in the above relationship can be described in the following manner. Driving is a visual task in which the peripheral vision plays a major role [18]. The quality of the useful visual field depends on several factors including the information processed in the peripheral area, foveal cognitive load, and age of the individual. A complex road scene (with road signs, obstacles, pedestrians, numerous vehicles, junctions, etc.) results in an increased spatial density, which ultimately decreases the useful visual field leading to a decreased driving performance. Speed, on the other hand, increases the amount of information to be processed per time unit, called the temporal density. According to Rogé et al. [18], there is no direct relationship between the speed and the driver's useful visual field; that is, the useful visual field deteriorates when the speed is increased. However, speed depends on the type of road (highway, city traffic, etc.) as well as drivers' adaptation to the road infrastructure by adjusting speed to minimize the effects of mental workload induced by the speed [11, 18]. Road infrastructure has implications to driver's vigilance as well; that is, monotony as a result of low sensory stimulation and low stimulus variation leads to decreased levels of arousal and alertness. Furthermore, the driver's useful visual field deteriorates with the prolongation of a monotonous task [18]. Therefore, it is important to consider the quality of the road scene (monotony) as well as the driver's state (vigilance) when taking into account the influences of speed on the driving performance.

2.3. Evaluating User Experience and Performance of Drivers. In general, the literature suggests three types of measures to evaluate the equivalence between driving in the real world and driving in a simulator: performance, physiological, and subjective measures [5, 11, 14, 17–20]. Performance measures evaluate physical and behavioral changes (e.g., vehicle speed, lane changing behavior, steering wheel variance, and head movements) and capture how well the user is performing a given task. Physiological changes (e.g., heart-rate variability (HRV), galvanic skin response (GSR), electrooculogram (EOG) signals, and electroencephalographic (EEG) signals) can capture a broad range of aspects of human cognition

and related processes. Although psychophysiological indices offer several advantages over other methods, they can bring confusions when interpreting the readings [21]. Finally, subjective measures are those that capture the user's subjective assessment of certain aspects using techniques like questionnaire and interviews. However, subjective measures are considered problematic because of the unreliability of self-reported emotional information and requirement to interrupt the experience [22].

Numerous studies have used one or a combination of the aforementioned measures when evaluating user experience and performance of drivers in any given context or for comparing those measures between different contexts. For example, Backlund et al. [1] report on a study that has evaluated a game-based driving simulator using questionnaire and interview data to capture opinions and attitudes from both students and instructors, and capture performance measures such as speed, headway distance, and lane change behavior. Another study [7] in which an eye tracker was used to detect distraction examined driver responses in a rear-end crash scenario during which the driver of the following car was distracted with a secondary task. Yet, another study [20] has estimated the driver's cognitive load based on the physiological pupillometric data (pupil diameter change) and driving performance data (variance of lane position and steering wheel angle). Nevertheless, these studies lack data enabling direct comparisons between driving in the real world and driving in a simulator and proper interpretations of results, which is something we wish to complement.

3. Method

3.1. Our Approach. Engen et al. [19] indicate three different types of environments in which traffic related experiments can be conducted along with their specific drawbacks: driving simulator, test track, and real traffic. Driving simulators lack the realism and the possibility to produce feelings of real danger. Test tracks lack the danger of interaction with other vehicles, and real traffic may be dangerous and hence not feasible for experiments. Moreover, it is unethical to expose subjects to risk in instrumented vehicles in potentially dangerous situations in test tracks or in real traffic. Since the aim of this paper is to evaluate the equivalence between driving in the real world and driving in a simulator, but taking into account the constraints and limitations of each of the driving environments mentioned before, we decided to involve only driving instructors in our experiment. The decision to involve driving instructors was also motivated by another reason, that is, to get an expertise perspective which minimizes the effects of noisy situational dispositions when involving humans in experiments as described thereafter.

Dreyfus [23] presents two concepts, based on Merleau-Ponty's Phenomenology of Perception, that are associated with intelligent behavior, learning, and skillful action, that is, the intentional arc and the tendency toward achieving a maximum grip. A skilled agent's skills are stored as more refined dispositions to respond to the solicitations of more and more refined perceptions of the current situation. As a result, the agent's body tends to respond to these solicitations in such

a way to bring the current situation closer to the agent's sense of an optimal gestalt, called the maximum grip. These allow experts, once immersed in the world of their skillful activities, not only to see what needs to be done, but also to do it intuitively and immediately. A study by Gilleade et al. [24] reports that novice players are more sensitive to challenges in game play than experienced players, which was observed in their physiological signals. However, age has a negative effect to the useful visual field, which may result in the driver neglecting some elements of information present in the road traffic [18]. Therefore, in identifying expertise we tried to limit our participants to middle aged drivers with a driving experience of at least about ten years. However, as we experienced difficulties in finding a large group of driving instructors alone in our experiment, we involved a group of regular drivers who have already obtained their driving licenses and had a similar amount of driving experience. Yet, this group did not get involved in real world driving as we found that it is difficult to continue the experiment in the real world environment involving a large number of individuals due to limited resources. Therefore, in our analysis, we tried to justify our selection based on the similarities of the measurements of driving instructors and regular drivers in the simulator context before proceeding to further analysis.

The simulator we involved in our study is a mid-range driving simulator (see Section 3.3 for details). At the beginning of our study, we assumed that the driver who drives in the simulator (also called the proband) will behave in at least a very similar manner as if he/she is driving a real car. However, this may not be the case as it depends on how well it can imitate the reality along with scenarios and physical behavior. Many advanced driving simulators are built satisfying these requirements to higher degrees (see [25]), but they are extremely expensive. However, the mid-range simulator used in the experiment can imitate scenarios and physical behavior to a satisfying degree. Since modeling of road scenes and reaching high fidelity are out of the scope of our study, we used thematically similar road scenes to mimic the real world circumstances, that is, highway and city traffic. Finally, we carefully planned the real world driving sessions during minimal traffic conditions of the day to make the real world and simulator traffic conditions approximately similar. Since we analyzed the driving behavior at two levels and considered situations of which the temporal resolution is low rather than instantaneous events, we deem that the difference between the road scenes is not substantial given that the traffic conditions are approximately similar. Indeed, the aim of our study is to evaluate the equivalence between driving in the real world and driving in a simulator already knowing that the two driving contexts are different in a number of ways while still being similar in theme and purpose. Moreover, as discussed in Section 2.1, there are implications of user experience in a carefully designed simulator for real world problem solving as it facilitates model-based thinking and learning beyond the limitations of the simulator promoted by a greater degree of self-efficacy.

In Section 2.3 we discussed different types of measures to evaluate the equivalence between different types of driving contexts along with their specific drawbacks. Our approach

is different from others in such a way that we first evaluated the driving behavior in two contexts based on performance measures alone and extended our analysis to involve psychophysiological measures. In other words, we analyzed a selected set of performance measures (e.g., speed and steering wheel movement) and we tried to infer possible motivations behind the variations of those measures based on psychophysiological measures (i.e., EEG based vigilance estimators). This type of analysis was possible in our study as we were interested in situations in which the temporal resolution is low; for example a driver may decide to drive faster in straight road segments, rather than instantaneous events, such as a driver looking in the side mirrors. Our approach allows us not only to evaluate the equivalence between the two contexts on drivers' consciously decided and physically observable actions, but also on unconscious and unseen influences on those actions. Another strong side of our approach is that we analyzed the data considering both grouping effects as well as individual differences.

3.2. Participants. A total of 14 healthy participants (mean age = 39.1 years and SD = 10.2 years; eight males and six females) took part in the experiment after providing a written statement of informed consent. The participants were recruited within two driver categories: driving instructors from a well known driving school (27–56 years; mean age = 40.9 years and SD = 11.5 years; five males and three females) and regular drivers within the university staff (26–51 years; mean age = 36.7 years and SD = 8.4 years; three males and three females). The driving instructors had been recruited by the driving school after considering their outstanding driving performance and number of years of driving (mean driving years = 23.6 and SD = 13.6). The regular drivers had comparative driving experience (mean driving years = 18.2 and SD = 8.3). After each experiment, each participant received a free lunch and refreshments as compensation for their involvement in the experiment.

3.3. Equipment and Tools. The experiment involved the mid-range driving simulator in the University of Skövde, Sweden [1]. It uses a real car with automatic transmission as a game control surrounded by seven screens. The screens cover the whole field of view for the driver, including the parts covered by the rear-view mirrors (220 × 30 degrees forward and 60 × 30 degrees rear). The physical feedback is comprised of sound vibrations and the car's fan also helps to create an illusion of movement. The sceneries and relevant physical behavior are generated by two different game engines: VDrift (<http://vdrift.net/>) and OGRE (<http://www.ogre3d.org/>). VDrift is a free and open source driving simulator in which the physical behavior is mainly inspired by Vamos automotive simulation framework (<http://vamos.sourceforge.net/>). OGRE is a scene-oriented, flexible 3D engine from which sceneries can be generated by integrating physical behavior using physics wrappers. Numerous studies have been conducted to evaluate the simulator's feasibility as a learning tool [1] and fitness for providing a higher sense of self-efficacy [6].

In our experiment we captured physical performance data in the following way. For the real car we used three linear string potentiometers (<http://www.advantagemotorsports.com/>) which we attached to brake and gas pedals and to the steering shaft. The status of each pot was sampled 20 times a second by an ATmega16 microcontroller which was recognized by the PC as a USB joystick. The speed of the car was derived based on readings from a GPS sensor. In the simulator, the car was equipped with two linear slider potentiometers attached to the brake and gas pedals and sampled at a rate of 100 times a second by an ATmega32 microcontroller in order to feed to the PC. However, for the steering we used a tooth wheel from a ball mouse which was attached to the steering rod near the front left wheel of the car, which has been already designed to automatically strive to have the wheels in a straight position when no one is driving the car. The rotation of the tooth wheel was read by an ATmega16 microcontroller and fed to the PC. Although there are certain differences in the sensors we used for capturing data in the two cars, we normalized the values of those sensors before sending the data to the recording software. For instance, in both cars, the reading of the gas pedal is zero when the gas pedal is at rest and one when it is fully pushed. The speed of the car in the simulator was obtained from the game engine itself. In addition to physical performance data, two cameras provided the frontal field of view and view of the subject in both cars.

Although most features of the two cars were similar, they had different transmission systems, that is, the real car had manual transmission where the car in the simulator had automatic transmission. We consider this to have minimal effect on the experimental conditions as driving instructors are used to both types of transmission systems and the specific variables we considered for the analysis (e.g., car speed and steering wheel movement) are to a greater degree independent of the type of transmission. Apart from the above sensors, driver's physiological data were captured using a low-cost sports heart rate monitor with chest belt—the Polar WearLink + transmitter with Bluetooth (<http://www.polar.fi/>) and Emotiv EPOC neurofeedback headset (<http://www.emotiv.com/>). The Emotiv EPOC headset is a low-cost alternative to highly expensive clinical type EEG equipment, but it uses 14 sensors and two references to capture EEG potentials following international 10–20 locations as well as providing two-axis gyro data for detecting head movements, namely, gyro x and gyro y . It has several other benefits such as wireless data transmission and being easy to setup. However, it has limitations as well, such as that it does not cover some important scalp positions, high signal-to-noise ratio, and lower sampling rate. Nevertheless, much research [26–28] reports the successful use of Emotiv EPOC neurofeedback headset to capture EEG signals for research purposes.

Captured data (vehicle dynamics, environmental conditions, and subject's physiological signals) were saved at their corresponding capturing points (personal computers) as well as in a central point as an effort to minimize risk of loss of data and synchronization errors.

The analysis was primarily carried out in Matlab [29] and graphs were obtained using Microsoft Excel.

Electroencephalographic (EEG) data was analyzed involving both Matlab and EEGLAB [30], which is an interactive Matlab toolbox for processing continuous and event-related EEG, EMG, and other electrophysiological data using independent component analysis (ICA), time/frequency analysis (TFA), and other methods. For comparing means of different groups, a balanced one-way ANOVA (Analysis of Variance) was used which is also available in Matlab as a function. ANOVA offers a greater flexibility for comparing means of even more than two groups which is not possible with Student's t -test [31, page 115].

3.4. Data Collection during Driving Tasks. Each driving instructor participated in the real world driving session and two driving sessions in the simulator, whereas each regular driver participated in two driving sessions in the simulator only. The real world driving session was approximately 20 minutes long in which the subject first drove on a road in city traffic until he/she reached a highway, next drove in the highway for several minutes, and finally drove back using the same route. In the simulator, each subject drove in the OGRE-based highway traffic track having levels of increasing difficulties for about 10 minutes, in VDrift Monaco track (city area like track, but no traffic) for about 5 minutes, and finally in VDrift LeMans track (landscape like track, but no traffic) for about 5 minutes. Figure 1 shows screenshots of the two driving environments and the three tracks of the simulator driving session. Table 1 shows the naming convention used in naming the tracks of each driving session and number of participants within each category of drivers who took part in driving on those tracks.

Each subject completed a questionnaire in a quiet office soon after each session of driving. In the questionnaire each subject had to answer questions about their driving experience, disturbances, and several other aspects, most of which were in 5-point Likert-type scale where 0 is not at all and 4 is extremely. However, only the question about disturbances from different sources of the questionnaire was considered in the analysis of this study.

3.5. Analysis of Performance Data. As indicated earlier, we captured four types of performance measures, that is, speed of the car, steering wheel angle, gas pressure, and brake pressure. The speed of the car was recorded in km/h whereas the other three action variables were converted into the normalized scale [0, 1]. Subsequently, the captured data were preprocessed to fix discontinuities and to synchronize between different data streams. Finally, segments of data have been identified based on the driving track boundaries noted in the corresponding video recordings. Figure 2 shows the variation of the four variables during one of the tracks.

Based on the four types of performance measures identified, we predicted a set of eight variables for further analysis: means of speed, means of steer, means of gas, means of brake, SDs of speed, SDs of steer, SDs of gas, and SDs of brake. This decision was partially motivated by the literature that suggests the use of steering wheel movement (SWM) for estimating the alertness level of drivers (e.g., [11]). The values for the above variables were calculated in the following way

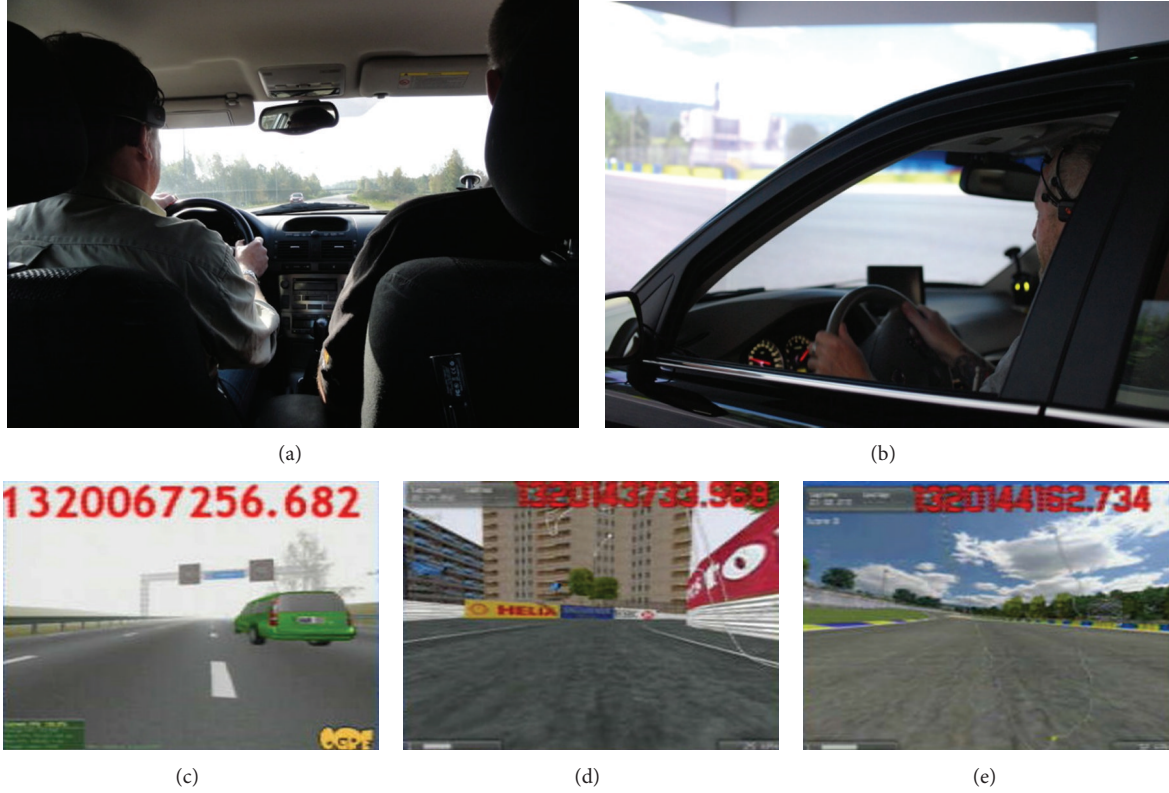


FIGURE 1: Screenshots of (a) the real world driving; (b) driving in the simulator; (c) OGRE-based highway traffic track; (d) VDrift Monaco track; and (e) VDrift LeMans track.

TABLE 1: Naming convention used in labeling different driving tasks.

Track	Description	Driving instructors	Regular drivers
RW.Tr.11	City traffic driving of onwards trip in the real world	8	0
RW.Tr.21	Highway driving of onwards trip in the real world	8	0
RW.Tr.22	Highway driving of the return trip in the real world	8	0
RW.Tr.13	City traffic driving of return trip in the real world	8	0
Sim1.Tr.30	Driving in the simulator in the in the highway traffic track during the first session	5	6
Sim1.Tr.41	Driving in the simulator in the VDrift Monaco track during the first session	7	4
Sim1.Tr.51	Driving in the simulator in the VDrift LeMans track during the first session	6	5
Sim2.Tr.30	Driving in the simulator in the in the highway traffic track during the second session	6	6
Sim2.Tr.41	Driving in the simulator in the VDrift Monaco track during the second session	6	5
Sim2.Tr.51	Driving in the simulator in the VDrift LeMans track during the second session	7	6

for instance, means of speed of a driver group is calculated by averaging each member driver's mean speed values of a given driving track where as SDs of speed of a driver group is calculated by averaging each member driver's standard deviation of speed values of a given driving track.

3.6. Analysis of Electroencephalographic (EEG) Data. The literature suggests different ways to analyze and interpret EEG data such as event related potentials (ERP) and power spectra analysis [32, 33]. For instance, theta rhythms intermittent in the band 6 to 7 Hz of $<15 \mu V$ in the frontal and frontocentral head regions are believed to be facilitated by emotions,

focused concentration, and during mental tasks [33]. However, we were unable to conduct our analysis based on the above frequently used techniques because our equipment had limitations caused by various artifacts, that is, higher amplitude and different shaped signals caused by sources such as body movements, eye-movements, impedance fluctuation, cable movements [32], and synchronization errors. Although the literature indicates certain techniques such as Independent Component Analysis (ICA) [30, 34] for removing artifacts, we could not succeed in our preliminary attempts with using ICA due to lesser number of EEG channels in the equipment. Therefore, we decided to use a different technique

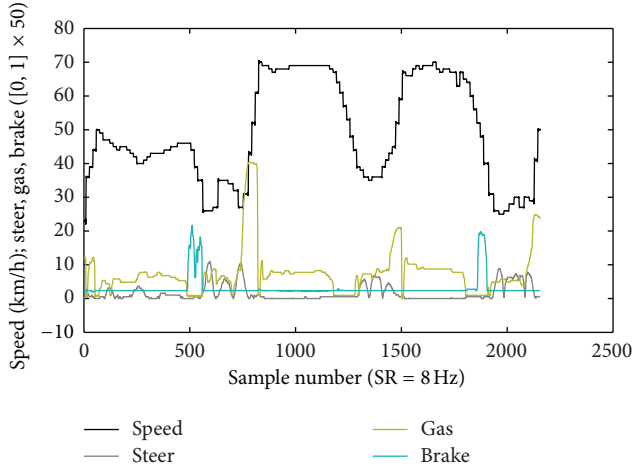


FIGURE 2: Variation of the four performance measures during one of the tracks (Tr.11) of a subject. Please note that the values of steer, gas, and brake were magnified 50 times for the illustration purpose.

as described thereafter. This method was partially motivated by the literature which suggests the use of EEG features with minute-scale smoothing for deriving vigilance estimators [14, 35, 36].

First, for each individual EEG recording, which consists of 14 channels of EEG data, we obtained the band powers for each of the seven frequency bands, that is, delta (1–4 Hz), theta (4–7 Hz), alpha1 (7–10 Hz), alpha2 (10–13 Hz), beta1 (13–22 Hz), beta2 (22–30 Hz), and gamma (30–45 Hz), which ultimately resulted in 98 (i.e., 14×7) band power components per each recording. Furthermore, for the aforementioned calculation, we involved the EEGLAB and the fast Fourier transformation (FFT) algorithm, and the band powers were calculated for consecutive one second durations of each component. However, as the band powers still contain noisy segments due to artifacts in the original EEG channels, we used the following technique to eliminate errors. First, in the respective EEG data channel of which the band power component was obtained, as well as in the gyro x and gyro y channels of that recording, signal magnitudes that exceeded the $3 \times \sigma$ (i.e., $3 \times$ standard deviation) level were identified. Next, a particular segment of the respective band power component was labeled as bad if at least one of the channels (within the respective band power channel or gyro x and gyro y channels) had already reported it as bad (i.e., that exceeded the $3 \times \sigma$ boundary). Finally, noisy segments were replaced using interpolation, which is based on adjacent band power values. After eliminating bad segments from band power components, each component was smoothed using the LOESS algorithm (local regression using weighted linear least squares and a second degree polynomial model) available in Matlab with approximately 30 seconds of time span. These band power components were treated as vigilance estimators, with the exception that only certain components can be associated with actual vigilance of drivers. Figure 3 shows a vigilance estimator derived from an EEG recording.

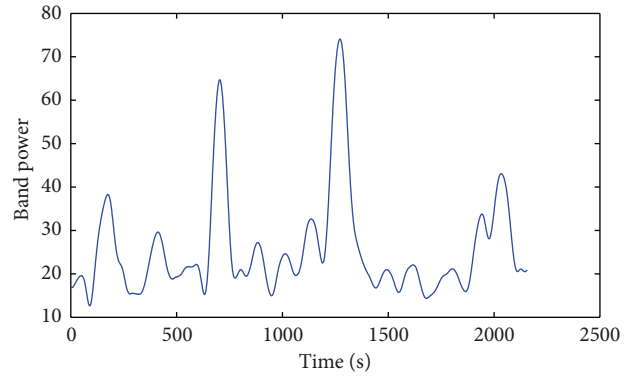


FIGURE 3: EEG-based vigilance estimator (derived from F3-Delta).

Once EEG features (i.e., vigilance estimators) were extracted from each recording, as the next step, each EEG feature was processed to find up to six peaks, which are highest, and up to six valleys, which are lowest, within each feature waveform. Next, of each peak and valley of a given feature waveform, the corresponding values of a given driving variable (i.e., speed, steer, gas, and brake) were obtained. However, before extracting the values, driving variables were smoothed to match a similar degree of smoothing span as of EEG features which is about 30 seconds. Smoothing of data has helped to leave out noise or other rapid changes in data.

After obtaining driving variable values at peaks and valleys of EEG features, ANOVA F -tests were performed to check whether means are different between a particular driving variable's values at peaks and valleys. For instance, ANOVA F -test for comparing the mean speed of peaks (44.6 km/h) and mean speed of valleys (51.2 km/h) at O2-beta of Tr.11 yielded that the means are significantly different (P value = 0.03). As the next step, the average values of peaks and valleys were obtained for each frequency band by considering only those values of which the means are significantly different (i.e., $P < 0.05$) between peaks and valleys. For instance, the above calculation has yielded the mean speed values 43.2 km/h and 50.9 km/h, respectively, for peaks and valleys of the delta band of Tr.11. Since we have not observed much difference between the values we obtained for each frequency band of a given track (e.g., $SD = 1.7$ and 0.8 , respectively, of the peaks and valleys of Tr.11), we obtained the averages of the values. For instance, the mean speed values of peaks and valleys are 42.7 km/h and 51.3 km/h, respectively, for Tr.11. We used these values when associating with corresponding performance measures (see Section 4.2).

4. Results

As we have already discussed in Section 3.5, we have predicted eight variables based on the four performance measures, that is, means of speed, means of steer, means of gas, means of brake, SDs of speed, SDs of steer, SDs of gas, and SDs of brake. However, most effective variables have to be recognized within those variables as not all variables are equally important when differentiating between the driving

TABLE 2: Multiway ANOVA F -test values for testing the effects of multiple factors.

Variable	Driver type $F(1,68)$	Driving session $F(1,68)$	Driving track $F(2,68)$
Means of speed	0.96	2.37	13.6***
Means of steer	0.99	0.92	24.7***
Means of gas	2.62	1.06	101***
Means of brake	0.01	0.35	1.04
SDs of speed	2.10	2.52	4.85*
SDs of steer	1.77	0.20	14.6***
SDs of gas	0.76	1.85	3.71*
SDs of brake	0.12	0.71	1.38

*, **, *** Significant differences at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

behaviors of the two contexts. However, as the first step, it has been required to check whether the driving behavior of driving instructors and regular drivers can be considered as similar or not, so that, if similar, we get eight participants for the real world driving and 14 participants for the simulator driving. ANOVA F -tests for comparing ages and experience between the two groups have showed that there is no significant difference between the ages ($F(1, 12) = 0.57$; $P = 0.46$) or experience ($F(1, 11) = 0.71$; $P = 0.42$). Table 2 shows multiway ANOVA F -test values for comparing means of each driving variable for testing the effects of driver type, driving session, and driving track. The test was performed considering only the data of the simulator driving experiment because the conditions were similar for both types of drivers in the simulator.

According to Table 2, there is no significant main effect for driver type or driving session. Therefore, the two driver categories have been considered as one category (i.e., as licensed drivers) and the two sessions have been considered as one continuous session for further analysis. However, Table 2 reveals that there is a significant main effect for driving track (P -values < 0.05 , except two variables). Therefore, based on the P -values that are lowest (i.e., $P < 0.001$), means of speed, means of steer, means of gas, and SDs of steer have been recognized as the most effective variables for differentiating the driving behavior of different driving tracks. However, to clarify the finding further, we prepared the following graphs (Figure 4) for each identified variable.

According to Figure 4, the patterns of means of steer and SDs of steer are to a significant degree identical, whereas the other two variables, that is, means of speed and means of gas, have distinguishing patterns. Therefore, we decided to consider only SDs of steer for our further analysis while leaving means of steer out. Our decision is partially motivated by the literature which reports successful use of steering wheel movement as discussed in the Section 3.1. After knowing that the driver type or the driving session does not play a significant role, but the driving track, and identifying the most effective variables, we proceeded to compare driving behavior of the two contexts. However, before that, we compared how the subjects have perceived the two environments subjected

TABLE 3: Comparing disturbances from different sources in the two driving contexts.

Source of disturbance	Real world driving [0, 4] M (SD)	Driving in the simulator [0, 4] M (SD)
EEG headset	1.4 (0.9)	1 (1.1)
Heart rate sensing equipment	0.3 (0.5)	0.4 (0.8)
Video recording equipment	0.5 (0.8)	0.3 (0.5)
Other equipment	0.9 (1.1)	0.9 (1.3)
Presence of researchers	0.5 (0.8)	0.4 (0.7)
When people talk	0.1 (0.4)	0.4 (1.0)

to their differences from different sources of disturbances due to experimental conditions (Table 3).

According to Table 3, none of the considered sources of disturbances has significantly disturbed the subjects as the mean values and standard deviations are very low. Moreover, both environments seem to be similar as the values are very similar in the two contexts.

4.1. Comparing Driving Behavior of the Two Contexts Based on Performance Measures. Figure 5 contains three graphs representing the behavior of each identified variable over different driving tracks of the two contexts.

As it can be seen in Figure 5, in general, standard deviations are higher in the tracks of the simulator context than these of the real world context of all three variables. Moreover, values of means of gas and SDs of steer are higher in the tracks of the simulator context than these of the real world context. Since implications of the above analysis are not very clear, scatter plots were prepared between means of gas and means of speed as well as SDs of steer and means of speed considering the values of each individual driver (Figure 6).

As it can be seen in both graphs of Figure 6, different clusters can be identified for each driving track except similar driving behaviors between Tr.11 and Tr.13 and between Tr.21 and Tr.22. Moreover, both graphs confirm that deviations (spread) of driving behaviors are higher in the tracks of the simulator context than these of the real world context. Apart from that, driving behavior of the real world context is attributed with lower values of means of gas and SDs of steering.

4.2. Comparing Driving Behavior of the Two Contexts Based on Both Performance and Psychophysiological Measures. Although psychophysiological data has been collected in two ways, that is, EEG and heart rate, we were unable to incorporate heart rate-based measures into the analysis because we observed abnormalities of data due to some technical problem in the equipment. For the analysis of EEG data we used a novel technique (see Section 3.6). Table 4 contains values obtained for peaks and valleys as well as mean values and standard deviations of each of the four performance measures and tracks of the two contexts. Figure 7 is a graphical representation of the values in Table 4.

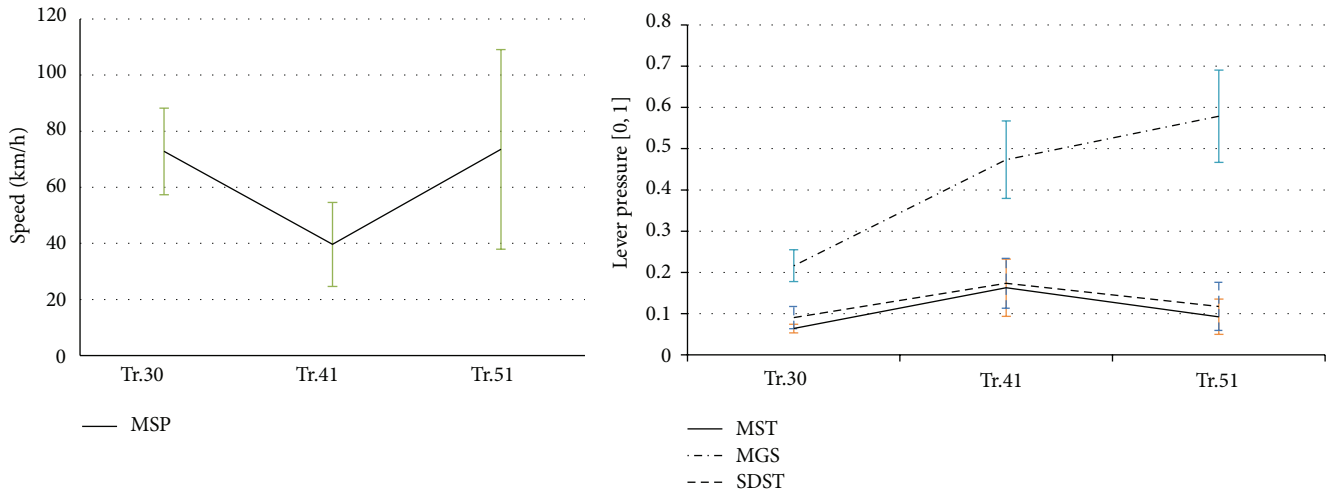


FIGURE 4: Values of means of speed (MSP), means of steer (MST), means of gas (MGS), and SDs of steer (SDST) in different tracks of driving in the simulator and standard deviations based error bars.

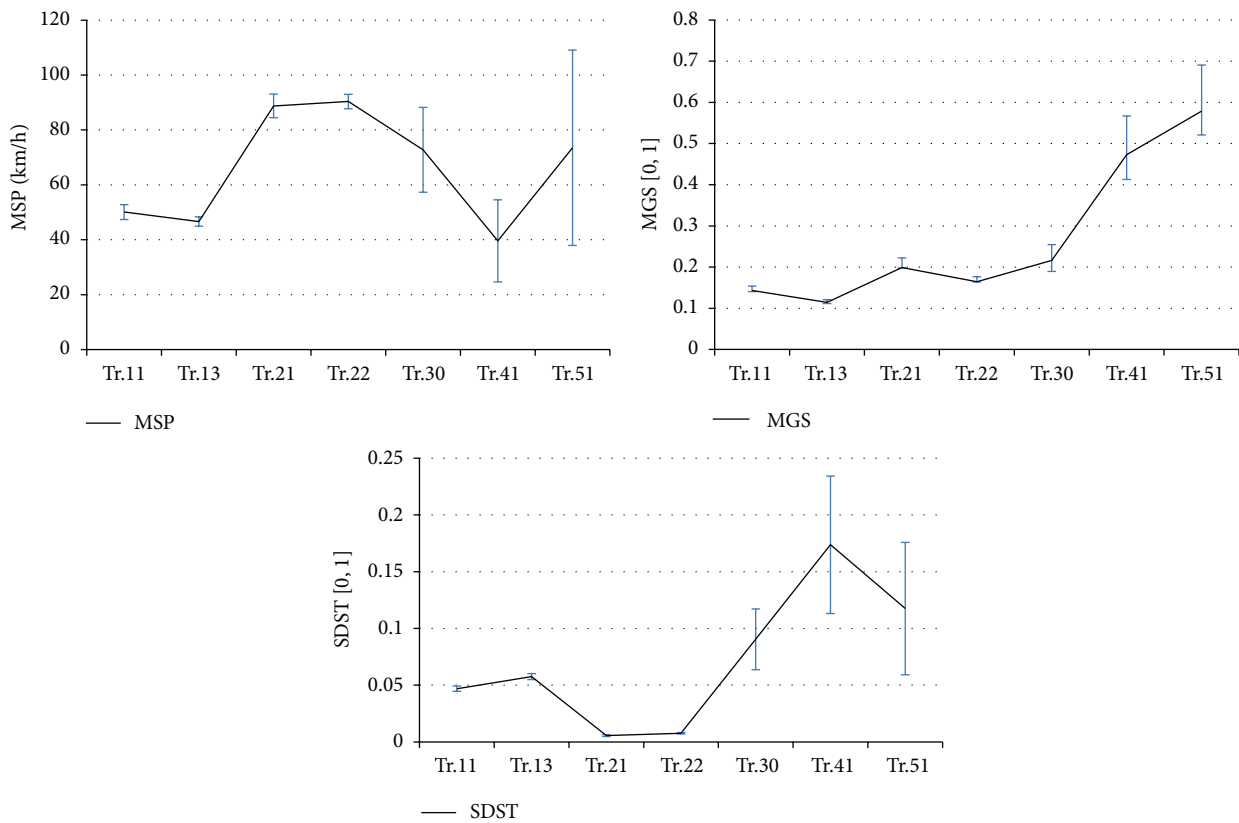


FIGURE 5: Values of means of speed (MSP), means of gas (MGS), and SDs of steer (SDST) in different tracks of the real world driving and driving in the simulator and standard deviations based error bars.

As it can be seen in Figure 7(a), means of speed (unconditioned) lies between the lines of mean speed at peaks and at valleys, and mean speed at valleys is above the two. Moreover, the distances between the lines of valleys and peaks are closer to each other, and means of speed is closer to mean speed at

valleys in the tracks of real world driving than these of the simulator.

Figure 7(b) shows the graph between means of steer (unconditioned) and mean steer at peaks and at valleys. Although the graph shows a similar pattern as of Figure 7(a),

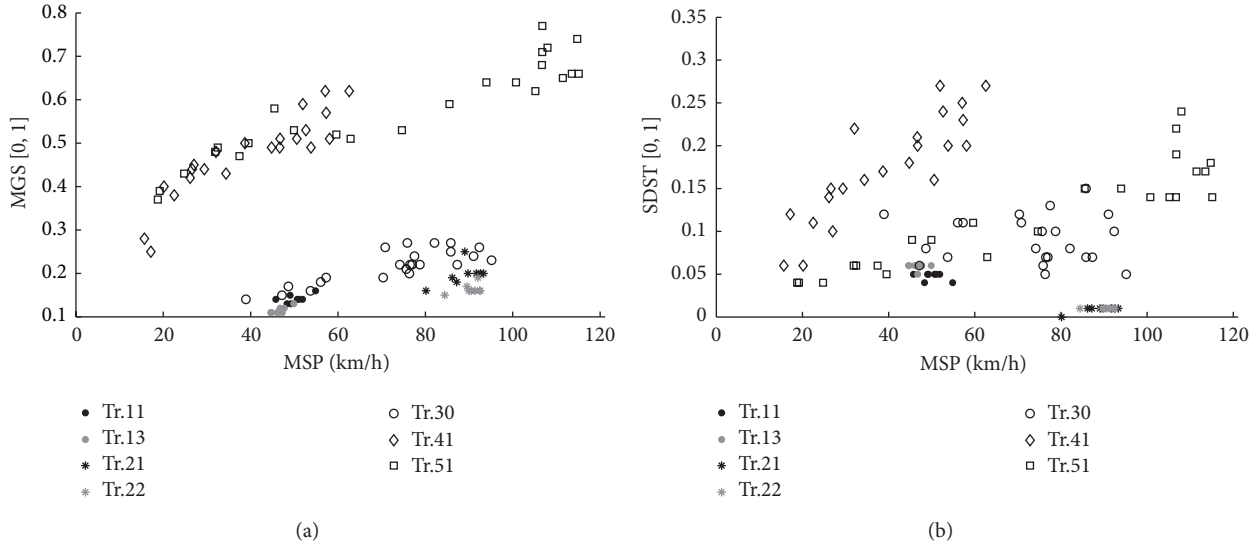


FIGURE 6: Scatter plots (a) between means of gas (MGS) and means of speed (MSP) and (b) between SDs of steer (SDST) and means of speed (MSP).

that is, means of steer lies between the lines of mean steer at peaks and at valleys, the line of peaks is above the line of valleys.

Although Figure 7(c) shows a similar pattern as of the other two, that is, the line of means of gas lies between the lines of peaks and valleys, the lines of peaks and valleys have crossed each other at certain driving tasks. When inspecting the behavior of these crossings, it can be seen that Tr.22 and Tr.30 driving tasks can be categorized as similar, both of which are associated with highway driving but in different contexts.

Finally, the lines of Figure 7(d) have a similar behavior as these of Figure 7(c) but an inverted behavior of the lines of peaks and valleys.

5. Discussion

In our study we evaluated the equivalence between driving in the real world and driving in a simulator at two levels, that is, using performance measures alone and by combining psychophysiological measures with performance measures. For the real world experiment, we involved eight driving instructors from a driving school and a car equipped with sensors to capture data about steering, gas and brake pressures, and speed. However, we involved eight additional drivers, who were regular drivers from the university, for the driving experiment in the simulator. Our analysis has shown that there is no significant difference (i.e., $P > 0.05$) between ages and experience of the two groups we involved, so we treated them as equal. The simulator too was equipped with sensors to capture similar data as of real world driving experiment. However, each participant participated in two sessions in the simulator but on two different occasions. Additional equipment was used to capture EEG data and heart rate of participants.

For the analysis, we predicted eight variables based on the four performance measures considered in the study, that is, means of speed, means of steer, means of gas, means of brake, SDs of speed, SDs of steer, SDs of gas, and SDs of brake. Based on the differences of the values of those variables in the simulator context, we were able to infer that the session does not have a significant influence on the driving behavior but driving track does. Moreover, the results confirmed that there is no difference between the two types of drivers we involved in our study. Further, we were able to identify means of speed, means of gas, and SDs of steer as the three most effective variables for differentiating the driving behavior of different driving tracks. So we used these findings as a basis when comparing the driving behavior between the two different contexts. A comparison between the possible sources of disturbances of the two experimental conditions, such as the mere presence of others, has revealed that both conditions are at least approximately similar.

In our analysis, in which we used only the performance measures, we found that the scatter plots between means of gas and means of speed as well as SDs of steer and means of speed are most effective when comparing the driving behavior of the real world and simulator contexts. In both scatter plots, the points representing different driving tasks (tracks) have been formed into different patterns of clusters. However, among those clusters, tasks of real world driving have a very low spread compared to the tasks of simulator which is again attributed with lower values of means of gas and SDs of steering. These results indicate an important aspect of simulator driving: people perceive the simulator as a more relaxed environment for experimenting with their skills, whereas in the real world they behave in a very restricted manner. Since this analysis does not reveal how people perceive the seriousness of their driving in the two contexts, we combined the psychophysiological features with performance measures in the second level of analysis.

TABLE 4: Values at peaks and valleys of EEG vigilance estimators, and means and standard deviations of the four performance measures.

Task	Mean speed at peaks (km/h)	Mean speed at valleys (km/h)	Mean speed (km/h)	Mean steer at peaks [0, 1]	Mean steer at valleys [0, 1]	Mean of steer [0, 1]	SD of means of steer [0, 1]	Mean gas at peaks [0, 1]	Mean gas at valleys [0, 1]	Mean gas [0, 1]	Means of gas [0, 1]	SD of means of gas [0, 1]	Mean brake at peaks [0, 1]	Mean brake at valleys [0, 1]	Mean of brake [0, 1]	SD of means of brake [0, 1]
RW.Tr.11	42.7	51.3	50.1	0.05	0.03	0.03	0.00	0.15	0.15	0.14	0.14	0.01	0.08	0.05	0.07	0.01
RW.Tr.13	36.5	47.6	46.7	0.06	0.03	0.04	0.00	0.09	0.11	0.11	0.11	0.01	0.14	0.09	0.10	0.02
RW.Tr.21	na	na	na	na	na	na	na	0.16	0.22	0.20	0.20	0.02	0.05	0.11	0.06	0.01
RW.Tr.22	84.0	90.4	90.4	0.01	0.01	0.01	0.00	0.19	0.13	0.16	0.16	0.01	na	na	na	na
Sim1.Tr.30	59.8	81.0	71.2	0.12	0.05	0.07	0.01	0.26	0.18	0.22	0.22	0.04	0.05	0.02	0.04	0.01
Sim2.Tr.30	61.0	80.6	74.2	0.09	0.05	0.06	0.01	0.24	0.19	0.21	0.21	0.04	0.04	0.02	0.02	0.01
Sim1.Tr.41	28.4	42.9	35.2	0.23	0.10	0.15	0.07	0.41	0.50	0.46	0.46	0.09	0.04	0.02	0.03	0.03
Sim2.Tr.41	36.9	51.6	44.0	0.24	0.12	0.18	0.07	0.43	0.55	0.49	0.49	0.10	0.06	0.02	0.04	0.03
Sim1.Tr.51	52.9	70.5	65.5	0.11	0.06	0.08	0.03	0.50	0.58	0.55	0.55	0.11	na	na	na	na
Sim2.Tr.51	62.6	91.7	80.4	0.17	0.07	0.10	0.05	0.51	0.61	0.60	0.60	0.12	0.06	0.02	0.03	0.02

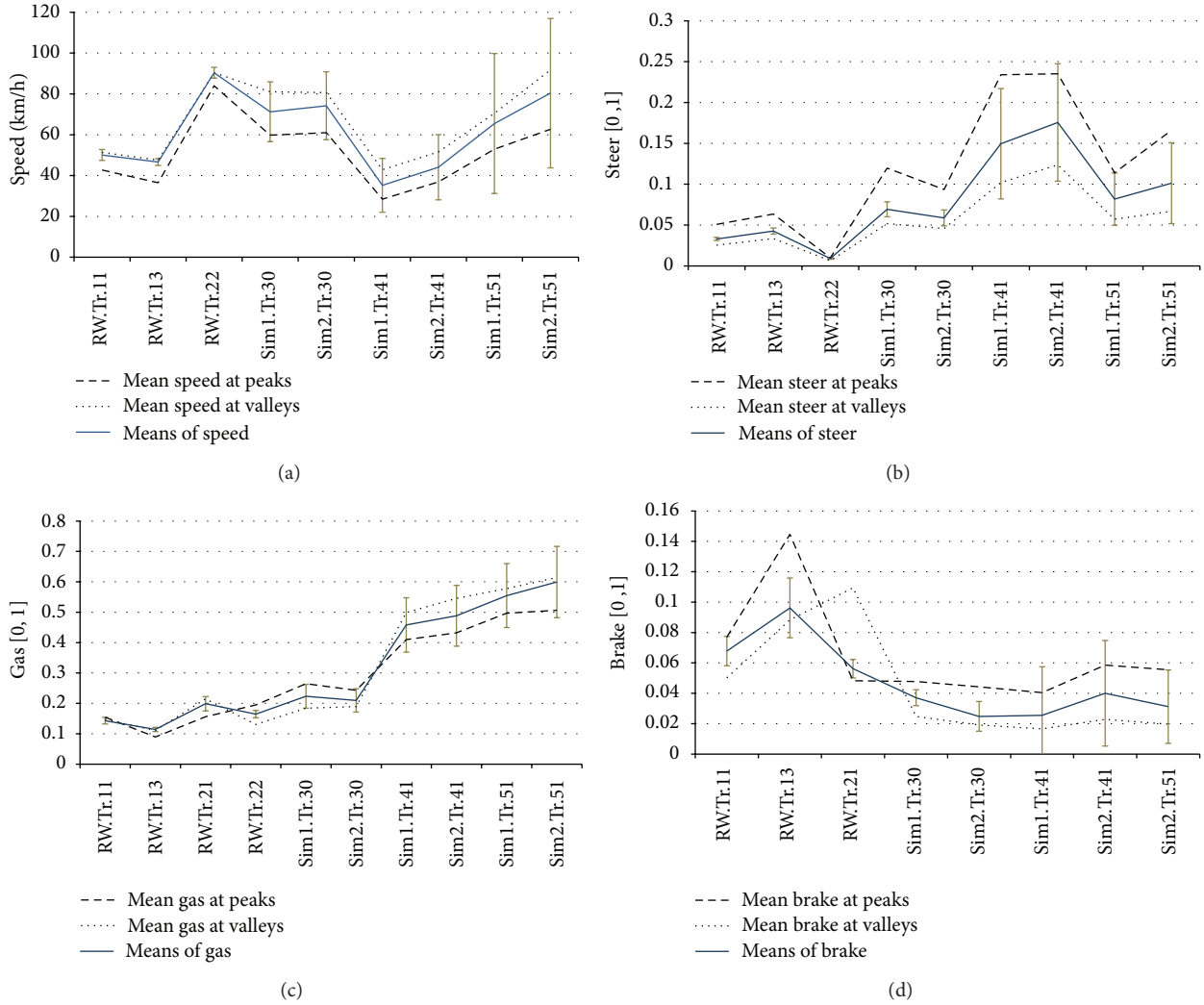


FIGURE 7: Graphical representations of the values at peaks and valleys of EEG vigilance estimators and means and standard deviations (as error bars) of the four performance measures: (a) speed, (b) steer, (c) gas, and (d) brake.

For analyzing EEG we used a method as indicated in the literature as a way to derive vigilance estimators from EEG data, but after improving its usefulness by associating its features with performance measures. This process has provided two values per variable, that is, values associated with high vigilance situations and values associated with low vigilance situations. The first result of this analysis, the graph between the mean speed values at peaks and at valleys and means of speed, indicates that drivers have maintained their mean speed within the limits of the speed levels that are associated with high and low vigilance levels. Moreover, it suggests that driving in low speeds is more vigilant than that at high speeds from the fact that the line of mean speed at peaks lies beneath the line of mean speed at valleys. Apart from that, driving in the simulator seems to be emotionally more relaxed than the real world driving as the distance between the lines of valleys and peaks is lesser in the tasks of simulator. However, drivers have tried to maintain their vigilance at a low level by driving in sufficiently high-speed levels in the real world which is

indicated by closer distances between the lines of means of speed and mean speed at valleys.

Although the graph between the mean steer values at peaks and at valleys and means of steer shows a similar pattern as does speed, the line of peaks is above the line of valleys, which suggests that higher degree of steering is associated with higher level of vigilance. Moreover, simulator driving tasks seem to be emotionally more relaxed than real world driving tasks as the distances between the lines of peaks and valleys of simulator driving tasks are less than these of real world driving tasks. Another observation is that the mean steering values are higher in the simulator than the values of real world driving, which may indicate the differences between the tracks of real world driving and driving in the simulator.

The graphs of the other two measures, gas and brake, show somewhat similar patterns; that is, the lines of peaks and valleys have crossed each other at certain driving tasks. The graph of the measure for gas indicates that accelerating

is more vigilant especially in the highway tracks of both contexts, but not in other tracks. This result can be explained in the following way: there is no need for a higher degree of gas on the highway, and accelerating can cause more stress as the speed increases. However, the graph of the measure for brake indicates that braking is less vigilant on the highway track of real world driving, while it is not on the other tracks of both contexts. This result can be explained as braking is required especially when there are disturbances such as other traffic and bends in the road which is true for all tracks except the highway track of real world driving. It is also observed that real world driving is attributed to a lower degree of gas and a higher degree of brake, whereas simulator driving is attributed to a higher degree of gas, except highway driving, and lower degree of brake. These results suggest the desire to drive in a relaxed mood in the simulator context.

Apart from the above findings, it is our belief that the zone between the lines of peaks and valleys, especially in the graph of speed, can be equated to the flow zone of Csikszentmihályi's flow theory, which states that strong involvement in a task (flow) occurs when the skills of an individual match the challenge of a task [22, 37, 38]. If our assumption is true, the challenging levels offered by certain tasks of real world driving are approximately similar to certain tasks of driving in the simulator. Further, the differences in the speed levels of different tasks can be explained in conjunction with the complexity of the road scene, that is, differences in the spatial and temporal density, as we have discussed in Section 2.2. For instance, speed level is higher on the highway track of real world than on the city traffic track. However, there are similarities of certain speed levels between the tasks of real world driving and driving in the simulator which may imply road scenes of similar complexities though we do not know what constitutes those complexities. Yet, as the individual differences of driving are higher in the simulator than the real world, we cannot justify the above implications with a greater confidence.

6. Conclusion

This paper has presented work comparing real world driving with driving in a mid-range driving simulator at two levels, that is, by using performance measures alone, by combining psychophysiological measures with performance measures, and by involving experienced drivers. The rationale behind involving experienced drivers was to get an expert perspective which equates with the evaluation of the simulator using a human driving model. Although it was not within the scope of the study to create models of high fidelity, we are confident that we have achieved a substantial progress with the aims of our study as the tasks of both contexts were thematically similar in most conditions.

In the first level of analysis, that is, analysis of performance measures, we were able to visualize the results in scatter plots which show distinguishing differences between the tasks of real world driving and driving in the simulator. For instance, clusters representing individual driving tasks of real world driving have a lower spread than the tasks of simulated driving. We equated this result to drivers' perception of the

simulator as a more relaxed environment for experimenting with their skills whereas the real world offers a very restricted driving environment. In the second level of analysis, which combined psychophysiological measures (i.e., EEG-based vigilance estimators) with performance measures, we were able to capture hidden physiological influences on drivers' performance in the two driving contexts. Results of this analysis further confirmed the findings of the previous level of analysis and helped to infer more findings.

Although there are certain limitations of our approach, such as low number of subjects and the fact that we did not involve novice drivers in the experiments, our belief is that both experimental procedures and findings of our experiment are very important to the field of serious games concerning how to evaluate the fitness of driving simulators and measure driving performance.

Acknowledgments

The authors wish to sincerely thank the staff of InGaMe Lab/Interaction Lab at the University of Skövde, staff members who voluntarily participated in the experiment, and Fästningens Trafikskola, Karlsborg. This work has been financed through the NeLC project of SPIDER program and the internal funding of the University of Skövde.

References

- [1] P. Backlund, H. Engström, M. Johannesson, and M. Lebram, "Games for traffic education: an experimental study of a game-based driving simulator," *Simulation & Gaming*, vol. 41, no. 2, pp. 145–169, 2010.
- [2] U. Ritterfeld, M. Cody, and P. Vorderer, "Serious games: explication of an oxymoron: introduction," in *Serious Games: Mechanics and Effects*, U. Ritterfeld, M. Cody, and P. Vorderer, Eds., pp. 3–9, Routledge, New York, NY, USA, 2009.
- [3] A. Brennecke, *A general framework for digital game-based training systems [Ph.D. dissertation]*, University of Rostock, 2009.
- [4] SWOV, "Simulators in driver training," SWOV Fact Sheet, Leidschendam, The Netherlands, 2010.
- [5] D. L. Strayer and A. Drewa, "Simulator training improves driver efficiency: transfer from the simulator to the real world," in *Proceedings of the 2nd International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, pp. 190–193, 2003.
- [6] P. Backlund, H. Engström, M. Johannesson, M. Lebram, and B. Sjöden, "Designing for self-efficacy in a game based simulator: an experimental study and its implications for serious games design," in *Proceedings of the International Conference Visualization (VIS '08)*, pp. 106–113, July 2008.
- [7] ITS, "Program Develops New Test Track Capability," The Sensor Newsletter—Winter 2004, ITS Institute, University of Minnesota, <http://www.its.umn.edu/Publications/Sensor/2004/Winter/Track.html>.
- [8] J. P. Gee, "Deep learning properties of good digital games: how far can they go?" in *Serious Games: Mechanics and Effects*, U. Ritterfeld, M. Cody, and P. Vorderer, Eds., pp. 67–82, Routledge, New York, NY, USA, 2009.
- [9] J. S. Warm, R. Parasuraman, and G. Matthews, "Vigilance requires hard mental work and is stressful," *Human Factors*, vol. 50, no. 3, pp. 433–441, 2008.

- [10] C. Collet, C. Petit, A. Priez, and A. Dittmar, "Stroop color-word test, arousal, electrodermal activity and performance in a critical driving situation," *Biological Psychology*, vol. 69, no. 2, pp. 195–203, 2005.
- [11] P. Thiffault and J. Bergeron, "Monotony of road environment and driver fatigue: a simulator study," *Accident Analysis and Prevention*, vol. 35, no. 3, pp. 381–391, 2003.
- [12] H. Ekanayake, P. Backlund, T. Ziemke, R. Ramberg, and K. Hewagamage, "Game interaction state graphs for evaluation of user engagement in explorative and experience-based training games," in *Proceedings of the International Conference on Advances in ICT for Emerging Regions (ICTer '10)*, pp. 40–44, September–October 2010.
- [13] D. Kahneman, *Attention and Effort*, Prentice Hall, Englewood Cliffs, NJ, USA, 1973.
- [14] C. T. Lin, R. C. Wu, S. F. Liang, W. H. Chao, Y. J. Chen, and T. P. Jung, "EEG-based drowsiness estimation for safety driving using independent component analysis," *IEEE Transactions on Circuits and Systems I*, vol. 52, no. 12, pp. 2726–2738, 2005.
- [15] E. Hudlicka, "Affective game engines: motivation and requirements," in *Proceedings of the 4th International Conference on the Foundations of Digital Games (ICFDG '09)*, pp. 299–306, Orlando, Fla, USA, April 2009.
- [16] M. D. Lewis, "Bridging emotion theory and neurobiology through dynamic systems modeling," *Behavioral and Brain Sciences*, vol. 28, no. 2, pp. 169–194, 2005.
- [17] D. Gruyer, A. Rakotonirainy, and J. Vignon, "Data combination using belief theory to assess driver's vigilance," in *Proceedings of the Australasian Road Safety Research, Policing and Education*, Wellington, New Zealand, November 2005.
- [18] J. Rogé, T. Pébayle, E. Lambilliotte, F. Spitzenstetter, D. Gisellebrecht, and A. Muzet, "Influence of age, speed and duration of monotonous driving task in traffic on the driver's useful visual field," *Vision Research*, vol. 44, no. 23, pp. 2737–2744, 2004.
- [19] T. Engen, L. Lervåg, and T. Moen, "Evaluation of IVIS/ADAS using driving simulators: comparing performance measures in different environments," in *Proceedings of the European Transport Conference*, The Netherlands, 2009.
- [20] O. Palinko, A. L. Kun, A. Shyrovkov, and P. Heeman, "Estimating cognitive load using remote eye tracking in a driving simulator," in *Proceedings of the Symposium on Eye-Tracking Research and Applications (ETRA '10)*, pp. 141–144, March 2010.
- [21] S. H. Fairclough, "Psychophysiological inference and physiological computer games," in *Proceedings of the Brain-Computer Interfaces and Games (Brainplay '07)*, Salzburg, Austria, 2007.
- [22] R. Picard, S. Papert, W. Bender et al., "Affective learning—a manifesto," *BT Technology Journal*, vol. 22, no. 4, pp. 253–269, 2004.
- [23] H. L. Dreyfus, "A phenomenology of skill acquisition as the basis for a merleau-pontian non-representationalist cognitive science," in *Proceedings of the Foundations and the Ontological Quest*, Rome, Italy, 2002.
- [24] K. M. Gilleade, A. Dix, and J. Allanson, "Affective videogames and modes of affective gaming: assist me, challenge me, emote me," in *Proceedings of the Conference of the Digital Games Research Association (DIGRA '05)*, pp. 547–554, Vancouver, Canada, 2005.
- [25] J. J. Slob, "State-of-the-Art driving simulators, a literature survey," DCT Report 2008.107, Eindhoven University of Technology, 2008.
- [26] A. T. Campbell, T. Choudhury, S. Hu et al., "NeuroPhone: brain-mobile phone interface using a wireless EEG headset," in *Proceedings of the 2nd ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld '10)*, pp. 3–8, New York, NY, USA, August 2010.
- [27] S. Wang, E. T. Esfahani, and V. Sundararajan, "Evaluation of SSVEP as passive feedback for improving the performance of brain machine interfaces," in *Proceedings of the ASME Design Engineering Technical Conference & Computers and Information in Engineering Conference (IDETC/CIE '12)*, Chicago, Ill, USA, August 2012.
- [28] H. Ekanayake, "Research Use of Emotiv EPOC," 2011, <http://neurofeedback.visaduma.info/emotivresearch.htm>.
- [29] MATLAB version 7.6.0. Natick, Massachusetts, Mass, USA, The MathWorks Inc., 2003.
- [30] A. Delorme, T. Fernsler, H. Serby, and S. Makeig, "EEGLAB Tutorial," University of San Diego California, 2006.
- [31] J. H. McDonald, *Handbook of Biological Statistics*, Sparky House Publishing, Baltimore, Md, USA, 2nd edition, 2009.
- [32] M. Teplan, "Fundamentals of EEG measurement," *Measurement Science Review*, vol. 2, no. 2, pp. 1–11, 2002.
- [33] W. O. Tatum, A. M. Husain, S. R. Benbadis, and P. W. Kaplan, *Handbook of EEG Interpretation*, Demos Medical Publishing, New York, NY, USA, 2008.
- [34] M. J. McKeown and R. Radtke, "Phasic and tonic coupling between EEG and EMG demonstrated with independent component analysis," *Journal of Clinical Neurophysiology*, vol. 18, no. 1, pp. 45–57, 2001.
- [35] D. Nie, X. W. Wang, L. C. Shi, and B. L. Lu, "EEG-based emotion recognition during watching movies," in *Proceedings of the 5th International IEEE/EMBS Conference on Neural Engineering (NER '11)*, pp. 667–670, May 2011.
- [36] L. C. Shi and B. L. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," in *Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '10)*, pp. 6587–6590, Buenos Aires, Argentina, September 2010.
- [37] A. L. Pedersen, *The correlation between self-reported game experience and biophysiological measurements [M.S. thesis]*, IT University of Copenhagen, 2009.
- [38] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," in *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era (MindTrek '08)*, pp. 13–17, Tampere, Finland, October 2008.

