A Missing Link between Fidelity and Realism: an experts’ assessment about an advanced motion-based driving simulator

Miguel Luzuriaga¹, Stefanie Trunzer², and Bernhard Schick³

¹,² University of Applied Sciences Kempten, Bahnhofstraße 61, 87435 Kempten, Germany
miguel.luzuriaga@hs-kempten.de
stefanie.trunzer@hs-kempten.de
³MdynamiX AG, Heßstraße 89, 80797 München, Germany
bernhard.schick@mdynamix.de

Abstract. A major concern about advanced motion-based simulators is their level of fidelity i.e., how close the motion sensation in a simulator is to the one perceived in a real vehicle. In this study, we collect the assessment from an exceptional sample composed by n=33 automotive industry experts who were asked to evaluate the fidelity in terms of steering, braking, and speed. Given the subjective nature of our measure, we propose a censored-data Tobit regression model that accounts for this issue, thus providing more accurate estimations. Our results show that, on average, experts evaluated the steering actions close to the maximum level of fidelity. However, braking and speed were evaluated lower in realism, and in fact, both diminished the overall fidelity judgement up to 50%. Moreover, coefficients indicate that steering contributes more to the judgment on fidelity than braking and speed actions. Heterogeneity in the experts’ responses and general implications are discussed.

Keywords: Driving Simulator · Fidelity · Real-world Driving · Motion-based Simulator · Validity · Subjective Driving Measures

1 Introduction

Testing the vehicles’ dynamics is critical for the automotive industry. It allows the development of new designs by e.g., doing iterations at early stages to gather reliable information about new concepts before they are fully commercialized. A key component to assess vehicle dynamics is motion feedback. Highly advanced motion-based driving simulators appear on this scene; however, their use has recently opened a vivid debate about whether they are reliable tools to carry out such tests. Perhaps the

* We would like to thank the Adrive Living Lab Staff for their assistance in implementing the experiment. In particular, we thank Stefan Vorderobermeier, Maximilian Böhle for their valuable assistance during the installation of the simulator, study setup, and comments.
main concern about motion feedback is the simulator’s capacity to achieve a reasonable level of fidelity on the perception of the driver i.e., a subjective evaluation about how close the motion sensation in a simulator is to the sensation in a real car. Moreover, the subjective nature of the driver’s perceptions makes it difficult to measure fidelity objectively. Then we ask, how can we contribute to solve this dilemma? Here we quote a phrase from Milton Friedman: “The only relevant test of the validity of a hypothesis is comparison of prediction with experience” The lack of empirical evidence from advanced motion-based simulators and the skepticism on the reliability of such evidence has motivated us to conduct a study that comprises the collection of data from an exceptional sample composed by automotive industry experts who were asked to test an advanced motion-based simulator. Overall, this study allows us to answer relevant questions such as, which are the determinants of experts’ perceived fidelity, to what extent those determinants contribute to their judgment, and what should be improved.

Besides the fact that our experiment generates evidence from a particular pool of subjects, we want to highlight other attributes from our study that differ from previous related work. First, we employ a highly Advanced Vehicle Dynamics Simulator (aVDS from ABDynamics) installed in our Adrive Living Lab from the University of Applied Sciences Kempten, Germany. This is a unique simulator in the region with a cost of approx. 3 million Euro, equipped with high responsive 6D motion, 210° cylindrical screen, and ultra-low latency (see Appendix A for more details). ABDynamics, the supplier of the simulator, is one of the world’s most recognized firms of automotive test systems. Second, we designed a 20-point scale to assess the perceptions about the simulators’ fidelity. This scale allows to obtain the maximum possible information which contrasts with well-known limitations from open questions, dichotomous questions, or few-points-scales. Next, we evaluate fidelity from three relevant variables: speed, steering, and braking. Speed, for instance, is considered among experts as the most reliable parameter to determine the validity of a simulator. The reason for this consideration, is that empirical evidence indicates that this is a parameter that has repeatedly demonstrated absolute validity i.e., the values obtained from a simulator are statistically not significantly different from those obtained from a real vehicle (e.g., Bella, 2008; Blaauw, 1982; Mueller, 2015; Reed & Green, 1999). Moreover, we employ steering and braking, both parameters repeatedly reported in previous research and associated to other relevant driving behaviors such as speed, lane position, and driving errors (e.g Reed & Green, 1999; Zöller et al., 2019). Finally, because the measure for fidelity is subjective and limited to an imposed maximum level from a scale, we employ a censored-data Tobit regression model that accounts for these issues. Most importantly, this model allows us to estimate our parameter of fidelity with higher level of accuracy than standard linear models do, such as the Ordinary Least Squares (OLS) The coefficients estimation indicates that, on average, experts evaluated the steering actions close to the maximum level of fidelity. However, braking and speed sensations were evaluated lower in realism, and in fact, both

---

1 Its clients include the top 25 global vehicle manufacturers, all seven Euro NCAP laboratories, and government test authorities.

2 Long (1997) provides evidence against the use of OLS, or even truncated regression due to lack of consistency of the parameter estimation.
diminished the overall fidelity judgement up to 50% with respect to steering. The magnitude of the coefficients also indicates that steering contributes more to the judgment on fidelity than braking and speed actions. Later in this paper, implications about these findings are discussed in more detail.

**Related studies:** Although in the common perception the use of simulators is recent, they have been used to study driving behaviors since the 1930s (Lauer, 1960). Since their appearance, the main concerns about their validity and reliability i.e., ability to accurately represent real-world driving and reproduce consistent results over time, remain until today (see Wynne et al., 2019 for a systematic review). Regardless these two properties are essential to conduct scientific-experimental research, and in turn used the evidence to draw meaningful conclusions, there is not a standardized methodology to assess their validity. Closely related to validity is the concept of fidelity, defined as the capacity to emulate real-world driving situations (Wynne et al., 2019). Given that validity is a broader term, typically used in research to describe for instance the internal and external validity, or the absolute and relative validity of research outcomes, throughout this paper we will refer fidelity as the capacity to emulate real-world situations. Regarding to this, Kaptein et al. (1996) define fidelity based on the simulator’s physical components. According to their classification, a highly valid simulator comprises three main physical components: large screens with a projected view close to 360°, a full-feedback motion base, and a vehicular cab with full controls. This classification, however, is only one among several fidelity measures such as the perceptual fidelity or the psychological fidelity (de Winter et al., 2007; Goode et al., 2013).

The rest of the paper is structured in the following way: in the next section, we describe the experimental design. In Section 3, we summarize our main results, and in the last chapter we present our conclusions.

## 2 Experimental Design and Procedure

The study was conducted at the Adrive Living Lab in Kempten, Germany. Our sample comprises *n* = 33 experts from the automotive industry. They include top managers, professors, and project engineers from 15 different firms and organizations. For information about the organizations and firms as well as about the subjects’ expertise, see Table 1 below. Subjects were invited by email and by a direct phone call. They were simply informed that the motion-based simulator was recently installed in our lab and that we would like them to get to know more about it and carry out an individual driving test. Next, each of them was assigned to a specific hour where part of our research team welcomed them and provided a brief introduction about the simulator. A member of the staff then showed them the hardware and asked them to get familiar with the projection, steering wheel, and pedals, as well as to adjust the seat and seatbelt (see Fig. 1 for details about the setup). The ride consisted in a simulated tour around typical street roads and highways and took an average time of 10 min. For the visualization and audio, rFpro software was employed, while for the control of the simulation, including a high-performance motion platform, the aVDS (advanced Vehicle Driving Simulator) software was used. After each ride, we provided to the par-
participants a fixed evaluation sheet based on a 20-point scale where 1= too weak [slow], 10= realistic, and 20= too hard [fast]. Below we present the wording of the questions:

**Speed:** “how was the feeling of speed during the ride?”
**Braking:** “how was the feeling of braking during the ride?”
**Steering:** “how was the feeling of steering during the ride?”

![Image](image_url)

**Figure 1.** Advanced Motion-Based Simulator and Experimental Setup

**Table 1.** Organization and Subjects’ Expertise

<table>
<thead>
<tr>
<th>Organization</th>
<th>Subjects’ Expertise</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADAC</td>
<td>Automotive Engineering</td>
</tr>
<tr>
<td>Applus IDIADA</td>
<td>Chassis Development &amp; Vehicle Dynamics</td>
</tr>
<tr>
<td>Audi</td>
<td>Calculation Engineering</td>
</tr>
<tr>
<td>Bosch</td>
<td>Product Innovation, Vehicle Sensors</td>
</tr>
<tr>
<td>Continental</td>
<td>Driving Simulation, Vehicle Testing &amp; Integration</td>
</tr>
<tr>
<td>Ergoneers</td>
<td>Vehicle Testing, Customer Care</td>
</tr>
<tr>
<td>Hyundai</td>
<td>Vehicle Testing &amp; Development</td>
</tr>
<tr>
<td>Hyundai HMETC</td>
<td>Driving Performance</td>
</tr>
<tr>
<td>Mercedes AMG</td>
<td>Calculation Engineering, Vehicle Dynamics</td>
</tr>
<tr>
<td>Porsche</td>
<td>Chassis Development &amp; Vehicle Dynamics</td>
</tr>
<tr>
<td>Porsche Engineering Services</td>
<td>Vehicle &amp; Structural Dynamics</td>
</tr>
<tr>
<td>Rimac Automobili</td>
<td>Vehicle Dynamics, Test &amp; Development</td>
</tr>
<tr>
<td>Schaeffler</td>
<td>Automotive Systems Engineering &amp; Vehicle Testing</td>
</tr>
<tr>
<td>Technical University Dresden</td>
<td>Driving Simulation</td>
</tr>
<tr>
<td>Volkswagen</td>
<td>Vehicle Dynamics</td>
</tr>
</tbody>
</table>
3 Results

In this section we summarize our main results. Below, we start by introducing basic descriptive statistics.

3.1 Descriptive Statistics

To study the distributional characteristics of our variables of interest, we present the following box plots (1-3) in Figure 2. The scores are based on the 1-20 scale where 10= maximum of fidelity, while 1 and 20 represent the extreme values i.e., “too slow/weak and too fast/hard”. Box (1) on the top, shows the scores from the judgment about the feeling of speed. Here, we see that the $Md=5$ and that the distribution is slightly skewed to the left. From the middle quartiles we also see that half of the experts’ valuations falls between the values of 3 and 7 suggesting that the speed in the simulator is considered slower than in real-world settings. In (2), we present the distribution of the scores about braking. Here, the distribution is also skewed to the right, however, the median $Md=6$ is slightly higher than the median in (1). In addition, we see from the middle quartiles marks that half of the scores are within the range of 4-9, indicating that in the experts’ opinion, the feeling of braking is slightly weak compared to real-world conditions. Finally, the box in (3) shows the distribution of scores related to the judgment of steering. Its $Md=10$ is the closest to the maximum level of fidelity, however, the distribution of the scores is very uneven, indicating that the judgment about the steering varies between weak, realistic, and hard. By observing the values between the middle and upper quartiles, we might infer that the steering is considered mostly realistic, but slightly hard. In comparison, most of the scores in (1) and (2) fall in the lower values of the scale, indicating that the speed and braking are considered, on average, slow and weak, respectively.

![Figure 2](image-url)

Figure 2. Distribution of Fidelity: (1) Speed, (2) Braking, and (3) Steering. Fidelity is measured in a 1-20 scale, where 10= the maximum of fidelity, while the extremes 1= too weak [slow], and 20= too hard [fast].
To further investigate the judgment on fidelity, we pooled the data from our three parameters: speed, braking, and steering into one parameter representing a unique measure of fidelity. In addition, we converted the scores from the 20-point scale into scores based on a 10-point scale. In this way, the value of 10 still represents the maximum level of fidelity, while 1 now represents the lowest value i.e., either too weak [slow] or too hard [fast]. The histogram in Fig. 3 shows the distribution of scores and the corresponding normal density curve for our constructed Fidelity parameter. Overall, the distribution indicates that the Fidelity of the simulator in terms of speed, braking, and steering is still considered as average from the experts. The curve for normality indicates a slight tendency towards the right tail, however, this might be due to the peak on the maximum level 10. The general opinions about the fidelity of the simulator are, however, heterogeneous. We see, for example, that the 2nd and 3rd peaks fall on lower values of the scale, 3 and 4, respectively.

![Figure 3](image)

**Figure 3.** Distribution of Fidelity and Normal Density Curve from the 10-point-scale where 10 is the maximum level.

Although boxplots and histograms are considered among the most helpful tools to uncover tendencies, from our data, we see that it is not easy to state conclusions with a reasonable degree of certainty. In the next section, we therefore apply formal statistical methods in order to achieve a higher precision in our estimations.

### 3.2 The Tobit Regression Model

To construct our regression model, we first transformed the scores from each of our three variables into a 10-point scale. For the purposes of interpretation, we codified Steering = 0, so that it becomes the reference group for comparison with the other two variables. Then we codify Braking = 1 and Speed = 2. Figure 4 shows the corresponding boxplots to describe the distribution of each variable. The box on the top shows the scores from the assessment on Steering (0). Here, we see an uneven distribution skewed to the upper quartile, and its median Md = 9.5 approaches to the maximum of fidelity. We also see that half of the sample falls between the values 7.5-10. Next, box (1) shows the distribution of the scores about Braking. Her we see a normal distribution with Md = 5.5 and where half of the observations falls around the values
suggesting that the assessment from the experts is somehow unified. In comparison, the distribution in (2) is slightly skewed to the lower quartile with $Md = 4$. Here, half of the scores fall between 3 and 6.

![Figure 4. Distribution of Fidelity: (0) Steering, (1) Braking, and (2) Braking. Fidelity is measured in a 1-10 scale, where 10= the maximum value, while 1= lowest value.]

To express the relationship between Fidelity and each of the simulation components, in Fig. 5 we estimate a regression line from the corresponding predicted values. From the graph we see that the feeling on the Steering was judged high on realism, however, this declines for Braking and Speed. More specifically, we see that judgment about the feeling of Braking marginally overcome the median, while Speed was assessed lower.

![Figure 5. Regression line for the relationship between Fidelity and each of the simulation’s components.]

Because the measure for Fidelity is subjective and limited to an imposed maximum level of 10, we propose a regression model that accounts for this issue and therefore applies a censoring on the data. In fact, when we look at the histogram in Fig. 3, we see that it is skewed towards the upper values, and in addition, it shows a peak at the highest value of 10. This suggests a censoring procedure from above i.e., when several cases with a value at or above this threshold takes place. In other words, the true value for Fidelity might be equal to 10 but it might be also higher than this value. The
same holds for a procedure involving censoring from below (from our lower threshold = 1). One might consider other methods to analyze this type of data, such as the standard linear regression OLS where 10 (1) will be treated as the maximum (minimum) actual value for Fidelity, and not as a upper (lower) limit of the maximum (minimum) fidelity. Long (1997) provides evidence against the use of OLS or even truncated regression due to lack of consistency of the parameter estimation. To achieve the highest possible level of consistency in our estimations, in Table 2 we present a Tobit regression model where data is censored at 1, as the lower limit, and 10 as the upper limit. The model aims to estimate the determinants of the simulator fidelity from three covariates treated independently, Steering, Braking, and Speed and has the following form:

\[ y^* = \alpha + \beta x + \epsilon \]

where:

- \( y^* \) is the non-observable Fidelity expressed as a latent variable
- \( \alpha \) = intercept
- \( \beta \) = coefficient for the Fidelity covariates
- \( x \) = categorical variable for Steering (reference= 0), Braking, and Speed
- \( \epsilon \) = error term

With respect to Steering (our baseline), the negative coefficients for Braking (-3.58) and Speed (-4.71) indicate that both significantly reduce the perceived Fidelity of the simulation with \( p < 0.01 \). However, by comparing the \( \beta \)s we see that it is Speed that diminishes the Fidelity to a larger extent (-4.71 vs -3.58). Why is this? Statistically speaking, Steering actions have more weight in achieving Fidelity than Braking and Speed. Our results on Speed contrast with most of previous studies using high-fidelity simulators which found that speed (or speed variation) was equivalent to real driving speed i.e., they found a significant correlation or non-significant difference between drives (see for example Abdel-Aty et al., 2006; Branzi et al., 2017; Fildes et al., 1997; Lee et al., 2013). Yet, in line with our results, several studies have reported non-valid speed comparisons (Carter & Laya, 1998; Fors et al., 2013; Hallvig et al., 2013; Santos et al., 2005; Senserrick et al., 2007; Törnros, 1998; Wang et al. 2010; Zöller et al., 2019). However, a more intuitive interpretation is that the simulation in terms of Braking and Speed needs to be improved because in the eyes of the experts these parameters have low levels of realism. Discrepancies between simulation and real-world driving were also found in Lee et al. (2002) and in Zöller et al. (2019). Furthermore, the intercept \( \alpha = 9.48 \) represents the Steering contribution to Fidelity, a value that is close to the maximum upper limit. In other words, if the simulation involved only Steering actions, the driving experience in the simulator would be assessed close to the maximum level of realism. Therefore, by including in the simulation parameters for Braking and Speed, the Fidelity drops up to 50%. In comparison, using a high and medium-fidelity simulators, Mueller (2015) and Reed & Green (1999) respectively, found a moderate positive correlation i.e., relative fidelity between steering behaviors in the simulator versus real-world driving. Finally, we controlled for motion sickness.
sensitivity by asking whether the participants felt sick during or after the ride. Our results show no evidence that this has affected the experts’ assessment.

**Table 2.** Determinants of Fidelity as a function of Steering, Braking, and Speed

<table>
<thead>
<tr>
<th>Regressor</th>
<th>β</th>
<th>Std. Error</th>
<th>t-value</th>
<th>p-value</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steering</td>
<td>(Reference parameter)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Braking</td>
<td>-3.58</td>
<td>0.824</td>
<td>-4.34</td>
<td>&lt;0.01</td>
<td>-5.22 -1.93</td>
</tr>
<tr>
<td>Speed</td>
<td>-4.72</td>
<td>0.825</td>
<td>-5.71</td>
<td>&lt;0.01</td>
<td>-6.36 -3.07</td>
</tr>
</tbody>
</table>

| Intercept | 9.48  | 0.785      | 12.07   | <0.01   | 7.91 -11.04 |
| Pseudo-R² | 0.078 |           |         |         |              |
| F-statistic | <0.01 |           |         |         |              |

Notes. Tobit regression models with robust std errors estimation clustered at the subject level.
Number of observations = 81.
Continuous variables measured from a 1-10 scale where 10= maximum of fidelity

4 Discussion

Regardless the general skepticism, there is somehow consensus about the advantages of the use of state-of-the-art simulators to evaluate vehicle dynamics, e.g., it allows to experience maneuvers that under real conditions would be implausible to conduct because of high economic costs, impractical logistics, important risks -such as those associated to the passengers physical integrity during the tests-, etc. On the other hand, their main drawbacks are considered those related to mimic real driving motion sensations. Without a reasonable level of fidelity, results from tests on vehicle dynamics have low validity and therefore becomes difficult to draw meaningful conclusions for the design of new concepts. In an attempt to contribute to the existing debate about the gap between what simulators can actually offer and real-world driving, we provide empirical evidence using a highly advanced motion-based simulator and collected from an exceptional sample composed by experts within the automotive industry. We let them evaluate the dynamics in the simulator in terms of three relevant driving parameters: steering, braking, and speed. Results from a Tobit regression model indicate that steering actions were evaluated close to the maximum level of fidelity. Braking and speed sensations, however, were assessed lower, and in fact, both tend to diminish the overall fidelity judgement up to 50% with respect to steering. These results exhibit important differences with respect to some of previous evidence. In terms of steering, while experts have evaluated it as high in fidelity, other studies found a moderate level (Mueller, 2015; Reed & Green, 1999). These discrepancies are more
noteworthy when comparing evidence about speed and braking. Several studies find these two parameters close to realism (e.g., Abdel-Aty et al., 2006; Branzi et al., 2017; Fildes et al., 1997; Lee et al., 2013), however, and in line with other fewer studies (Carter & Laya, 1998; Fors et al., 2013; Hallvig et al., 2013; Santos et al., 2005; Senserrick et al., 2007; Törnros, 1998; Wang et al. 2010; Zöller et al., 2019), we find the opposite. As we can see, we are still far from consensus both in terms of fidelity measurement methodologies, and about results. This should motivate new research on, for instance, reliability in measuring methods, differences between common drivers’ and experts’ assessment, or what other parameters moderate fidelity.

References


Biographical notes: Miguel Luzuriaga received a BBA in the Universidad de las Americas-Puebla, Mexico, a Master in Business and Economics, and PhD in Experimental Economics both in the University of Stavanger, Norway. He is currently working as Researcher at the Adrive Living Lab (Research Center Allgäu). His research interests are in Experimental and Behavioral Economics, particularly in advanced driver-assistance systems (ADAS), driver’s behavior, decision-making under risk, and statistical methods. He has published several papers in international peer-reviewed
journals and had referee assignments for the Scandinavian Journal of Economics and for the Review of Behavioral Economics. In addition, he has taken part in research and development projects for the automotive industry.

Stefanie Trunzer achieved a Diploma of Mechanical Engineering and a Master of Advanced Driver Assistance Systems at University of Applied Sciences Kempten. Currently she is working at Kempten University as a researcher for simulation of advanced driver assistance systems. She is responsible for managing the new driving simulator to make the driving functions work like in a real car and therefore be able to conduct interaction studies between drivers and assistance systems.

Bernhard Schick is the CEO of MdynamiX AG and Professor at the University of Applied Sciences Kempten in the field of Advanced Driver Assistance Systems, Automated Driving and Vehicle Dynamics. He is also the head and founder of the Adrive Living Lab which is part of the Research Center Allgäu. In addition, he has published several papers and has patents including methods for monitoring the tires, and vehicle diagnostic and sensor systems. He was Pro Superbike Team Champion (1993 and 1994) and International German Road Racing Master Supersport Bikes in 1996.