Is simulation (not) enough?
Results of a validation study of an autonomous emergency braking system on a test track and in a static driving simulator

Martin Jentsch & Angelika C. Bullinger
TU Chemnitz
Germany

Abstract

Comparison of data gathered with real vehicles and with a driving simulator is still heavily debated. This paper provides results of a validation study with 164 participants who tested an autonomous emergency braking system (AEBS) either in a driving simulator or on a test track. Participants were similar concerning age and driving experience and experienced real driving on a test track and in a 180° Field of View (FOV) static driving simulator. Study design, scenarios and questionnaires to assess e.g. drivers’ perceived degree of dangerousness of the situation, perceived usefulness of the system in each scenario and overall acceptance were used in both set ups. Additionally, vehicle dynamics were recorded. Participants drove one of six types (three braking intensities each with two different times for acoustical warnings) of the system. Three traffic scenarios (e.g. distracted driver with a sudden braking of the leading vehicle) with a moving vehicle ahead and two scenarios with a stationary target (e.g. AEBS intervention during evasive manoeuvre) were accomplished by each participant. It was found that participant’s reaction in the simulator is comparable to the reaction on the test track. Participants’ judgment of the system, situation and overall acceptance could be shown to be almost the same.

Introduction

In the last 20 years Advanced Driver Assistance Systems (ADAS) developed and diffused rapidly. Despite their benefits, it cannot be ignored that the driving task can seriously change if driving with ADAS support. One example is the additional need of the continuous monitoring of the ADAS (Spanner-Ulmer, 2008). For this reason, it is necessary to ensure that the driver quickly understands the function and the boundaries of the system and is able to operate it safely. This is necessary to ensure that the driver does not put himself or other road users at additional risk, for instance by misinterpreting the ADAS. If these prerequisites are fulfilled the driver is more likely to accept the system, is willing to use it continuously and the ADAS can reach its full potential to increase traffic safety or drivers’ comfort. When developing ADAS, the manufacturers have to face the challenge to design the systems according to the driver’s needs and to ensure technical and functional reliability. To determine suitable specifications of a new ADAS, requirements are usually obtained by studies

with participants, ideally the future customers of the company, who experience and evaluate the system in early development stages. These experiments can be performed in real road traffic as a “field” or “on road” test, on a test track or in a driving simulator.

Surprisingly, there is little knowledge to which extent results from experiments with ADAS are comparable between driving simulators and real vehicles. This refers on the one hand to possibly modified (driving) behaviour (objective measures of the vehicle dynamics and the drivers’ behaviour) in the driving simulator. On the other hand, the subject’s assessment and the overall acceptance (subjective measures) of the ADAS may be different in the driving simulator. Furthermore it is uncertain whether the relations between objective and subjective measures are influenced by the test environment.

This leads to the research questions:

- Is it possible to get similar findings concerning driving/driver’s behaviour, system evaluation/overall acceptance and situation evaluation in a static driving simulator compared to findings from a real vehicle for interventions of an autonomous emergency braking system (AEBS)?
- Which objective and subjective measures are suitable for ADAS evaluation in a static driving simulator?

If the feasibility of an experimental procedure in the static driving simulator can be demonstrated this will, of course, not entirely substitute tests with real vehicles. Functional reliability of the ADAS and a final subject assessment will always be necessary in a real vehicle. However, it would be possible to perform experiments in early concept or very early development phases of the ADAS without the need of a fully operative ADAS in a real vehicle. This allows important insights that are valuable for the design of the system, which can positively influence the development process. When optimization potentials regarding the ADAS specification are found as early as possible, development times can be shortened and development costs can be reduced.

In this paper, the validity of a static driving simulator for experiments with ADAS, which actively intervene in the longitudinal control of the car, will be examined. Therefore the ADAS Aktive Gefahrenbremsung, an Autonomous Emergency Braking System (AEBS), which was developed in the AKTIV research initiative, was selected as example. The AEBS enables autonomous prevention of rear-end collisions without driver’s action. Therefore it is representative for similar ADAS, which actively intervene in the driving task and systems, which will enable highly or fully automated driving in the future.

---

3 AKTIV was funded by the Federal Ministry of Economics and Technology (BMWi). TU Chemnitz did the evaluation of the Aktive Gefahrenbremsung on a test track as part of a subcontract.
State of the art

The driving task (Donges, 1982; Bubb, 2003) represents a highly complex task for the driver in which he may reach his limits concerning human perception and reaction. Traffic accidents can be the consequence of a non- or only bad performed completion of the driving task (Lee, 2008). Accidents in longitudinal traffic, such as rear-end collisions, account for a percentage of about 25 - 30% (Hannawald, 2013) of all traffic accidents. This illustrates the high potential for safety ADAS, especially in longitudinal control, such as AEBS. Particularly for actively intervening ADAS it is crucial to know and ensure the driver’s interaction with the vehicle and the ADAS. This is even more important for systems which intervene at higher speeds to prevent misuse, abuse and associated negative effects of the ADAS relating to road safety (Knapp et al., 2009).

Measuring driver and system performance

Characteristic values to evaluate the driver’s interaction with ADAS can be divided into physically measurable, objective measures and subjective measures, which are obtained by interviewing the driver, e.g. using questionnaires.

Physically measurable, objective measures can be subdivided into vehicle dynamics measures and driver behaviour measures (Wierwille et al., 1996; Johansson et al., 2004; Östlund et al., 2005; Dotzauer et al., 2011; Dettmann, 2012). From the recorded raw data further values such as minima, averages or maxima can be calculated within specified measurement intervals in order to derive results concerning the desired research question.

Subjective measures can be divided into measures of acceptance (Arndt & Engeln, 2008), system evaluation (Riedel & Arbinger, 1997) and situation evaluation (Kiefer, Flannagan & Jerome, 2006). For ADAS that are not on the market and therefore cannot have been experienced by drivers yet measuring acceptance is difficult. In the experiment drivers experience a new ADAS for the first time and only over a limited period of time. Referring to the theory of planned behaviour (Ajzen, 1991), the attitude toward the behaviour, the intention (to use the ADAS) and the perceived system’s characteristics can be good predictors of the future driver’s acceptance.

With questionnaires, it is possible to let the driver assess the perceived usefulness and usability (Fastenmeier & Gstalter, 2008) or the overall satisfaction (Pataki, 2005) of an ADAS. Since ADAS that focus on increasing traffic safety can only be experienced in complex or hazardous traffic situations, the situation evaluation is closely linked to the system evaluation. The perceived driving situation can be measured by participants’ perceived danger of the situation, the characteristics of the situation, e.g. concerning crucial object in the scenario or the estimation of distances. Measures of acceptance are giving developers insights about the driver’s attitude towards the ADAS and provide an estimation of his actual will to use the system in real traffic.
Based on these considerations a set of the most frequent and according to the literature most promising objective measures to assess systems’ and driver’s performance for AEBS was chosen and questionnaires were designed for the experiments (see table 1).

Table 1. Objective and subjective measures for the validation study

<table>
<thead>
<tr>
<th>Objective measures</th>
<th>Subjective measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vehicle dynamics</strong></td>
<td><strong>Acceptance</strong></td>
</tr>
<tr>
<td>• Speed</td>
<td>▪ Attitude toward the behaviour</td>
</tr>
<tr>
<td>• Longitudinal acceleration</td>
<td>▪ Intention (to use the ADAS)</td>
</tr>
<tr>
<td>• Distance (incl. Time Headway (THW) &amp; Time to Collision (TTC))</td>
<td>▪ Perceived system’s characteristics</td>
</tr>
<tr>
<td>• Brake Reaction Time (BRT)</td>
<td></td>
</tr>
<tr>
<td>• Pedal Measures / pedal activity</td>
<td></td>
</tr>
<tr>
<td>• Steering behaviour/ steering wheel angle</td>
<td></td>
</tr>
<tr>
<td><strong>Driver behaviour</strong></td>
<td><strong>Situation validation</strong></td>
</tr>
<tr>
<td>• Glance behaviour</td>
<td>▪ Danger</td>
</tr>
<tr>
<td></td>
<td>▪ Objects in a situation</td>
</tr>
<tr>
<td></td>
<td>▪ Distances in a situation</td>
</tr>
</tbody>
</table>

Issues regarding the driver-vehicle-interaction during the intervention of an AEBS and acceptance towards the system cannot be answered in real road traffic. The main reason for that is that these systems can only be experienced in perilous situations. This causes a far too big threat to the safety of the participant and other road users which disqualifies the test environment “real road” for AEBS experiments in early stages. Therefore, only the test environments test track and driving simulator are suitable.

**Test methods**

The three main test quality criteria objectivity, reliability and validity are the basic requirements while planning, conducting and interpreting experiments (Bryant, 2000). If the experiment does not take place in the real road traffic, questions concerning the validity of the findings may occur. Two kinds of validity can be distinguished: internal (adequately accurate acquisition of parameters, extent to which a causal conclusion based on a study is warranted) and external (transferability and generalizability of the results). For comparative studies between e.g. a driving simulator and a test track experiment, as conducted in the case at issue for this article, the external validity can be distinguished into two types (Blana, 1996): absolute validity (same or similar measured values between the test environments) relative validity (same effects or rank order but different absolute...
values, depending on the test condition between the test environments). By examining the above mentioned measures regarding external validity, assumptions can be made concerning the feasibility of the static driving simulator for AEBS or similar ADAS evaluation experiments.

**Empirical Study**

*Function and types of the AEBS*

Six different types of the AEBS have been developed by the *AKTIV* research initiative. All of these are designed to avoid accidents within the limits of the system and to brake the vehicle until standstill. The types consist of three braking intensities (full, partial and combined braking) with two different times for acoustical warnings each. The full braking (FB) intervenes until standstill with a deceleration of 7 m/s². In the partial braking (PB), the vehicle decelerates until standstill with 4 m/s².

![Figure 1. Types of the AEBS grouped by braking intensity and warning times](image)

The combined braking (CB) decelerates with 4 m/s² in the first 0.7 s and afterwards with 7 m/s² until standstill. For each strategy of braking there is an acoustic warning, which starts either simultaneously or 0.5 s before the braking intervention of the AEBS. Apart from the characteristics of each type the factors warning time and braking intensity can be compared. Figure 1 shows the types of the AEBS with their principal modes of operation.

The six types were implemented in two vehicles for the test track experiment. For the presentation of the AEBS in the static driving simulator they were simulated with the software “SILAB 3.0” given identical system characteristics to those in the real vehicles.
The airport and test area Großenhain was chosen as test track. The driving simulator experiment took place in a static driving simulator with 180° projection at the Technische Universität Chemitz (Jentsch, 2014).

Scenarios

Since the AEBS is an ADAS that is currently not available it is of great importance to illustrate the participants the functioning of the AEBS in various scenarios in a fixed order. Not only was a reliable assessment of the acceptance towards the system and the system evaluation by the participants in the focus of the experiment. Also knowledge about the interaction of the driver with the AEBS had to be examined. Therefore it was necessary to define different scenarios with respect to everyday situations where an AEBS would intervene.

Scenario 1 - Unexpected impending frontal collision with visual driver distraction: AEBS intervention to prevent a rear-end collision with a vehicle in front, which suddenly decelerates. A moving target, which decelerates unexpectedly for the participant while he is distracted by a visual-motoric secondary task and driving at 60 km/h, is used to represent the scenario.

Scenario 2 - Stop&Go situation without driver distraction: Scenario 2 represents a classic Stop&Go situation at low speed (maximum 40 km/h). This scenario represents an accident hotspot in longitudinal traffic (Schaller, 2009). It may also illustrate the participant the possibly existing disadvantages of the warning signal before the braking intervention.

Scenario 3 - Announced AEBS intervention without driver distraction: Scenario 3 represents a modification of Scenario 1, where this time the participant is not visually distracted. Additionally a verbal explanation of the AEBS function is given before starting the scenario. The participants should experience consciously the AEBS intervention to get a better understanding of the timing and intensity of braking and acoustical warning.

Scenario 4 - Unexpected AEBS intervention during evasive maneouvre: Depending on vehicles’ velocity and the intensity of deceleration during the braking manoeuvre it is possible that the distance necessary for fulfilling an evasive manoeuvre is shorter than the necessary distance for braking. Problems can occur for AEBS because the system may detect a critical situation and starts intervening. However the driver may plan an evasive or overtaking manoeuvre. Scenario 4 is used to examine this driving situation with a velocity of 65 km/h.

Scenario 5 - Announced AEBS intervention when approaching a stationary obstacle: At the end of the experiment the participants are asked to compare all six types of the system in scenario 5. For this purpose, the participants drive consciously and without visual distraction with a predetermined speed (50 km/h) towards a stationary target. Before the first run in scenario 5, the participants are explained that there are six types of the system without going into their characteristics. A comparative analysis of the subjective assessment of the system between the types of the system can be made with the help of this scenario.
**Study Design**

Up to scenario 3, a between-subjects-design was chosen and a moving target was used (see figure 2). Each participant consistently drove these scenarios with the same type of the AEBS. By doing so, each participant should get the opportunity to estimate each detail of the system at different speeds and in different situations. The system types were assigned to the participants in a way that every type was driven by the same number of male and female participants and occasional and frequent drivers.

*Figure 2. Moving (left) and stationary (right) target on the test track (Jentsch et al. 2012)*

For scenario 4 and 5, a stationary target was used (see figure 2). In scenario 4, the between-subjects design of the first three scenarios was generally maintained. Types of AEBS with full braking interventions were excluded from this scenario due to the short necessary distance for braking. Participants with FB types were equally assigned to partial or combined braking intensities in scenario 4. For scenario 5 a within-subjects design was chosen. Participants began with types of the AEBS they already were familiar with from the first three scenarios. This was followed by runs with the other five types. The order was balanced to eliminate position and sequence effects.

**Participants**

79 people joined the experiment on the test track, 92 took part in the driving simulator. Two participants of the 92 already dropped out during training sessions other five had to end their attendance due to simulator sickness (Reason, & Brand, 1975). In the end data of 85 participants was collected in the driving simulator. Participants were recruited from a database, with flyers and announcements. They were given a reward of 25 € for participating on the test track were the experiment lasted approx. 2.5 h and 15 € in the driving simulator where the duration was approx. 1.5 h.
To investigate the influence of driving frequency, groups of occasional (< 10,000 km/year) and frequent drivers (> 15,000 km/year) were formed. Sex of the participants was almost equally distributed within these groups to minimise the influence of the participant on the comparison of the experimental environments the samples were taken such as participants were similar concerning age and driving frequency in both test environments. Table 2 shows participant’s characteristics on the test track and in the driving simulator.

Table 2. Age of participants and kilometres driven within the last year

<table>
<thead>
<tr>
<th>Participant’s characteristics</th>
<th>Test track</th>
<th>Driving simulation</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age [years]</td>
<td>79 28.9</td>
<td>90 28.7</td>
<td>167 0.128 .898</td>
</tr>
<tr>
<td>km driven last year</td>
<td>79 13678</td>
<td>90 14471</td>
<td>167 -0.455 .650</td>
</tr>
</tbody>
</table>

Data analysis

To answer the questions mentioned in the introduction it is necessary to examine, to what extent a dependency of the (driving) behaviour, the subjective assessment and the relations between the (driving) behaviour and the subjective assessment of the types of the AEBS exists in both test environments. Furthermore the results must be analysed in the context of different driving scenarios and whether there are differences in the (driving) behaviour and the subjective evaluation of the AEBS depending on the annual kilometres covered by the driver. Figure 4 summarises these issues and illustrates the dependent and independent variables for the validation study.
In order to make a reliable conclusion about the suitability of a static driving simulator for the investigation of an AEBS, it is necessary to measure the dependent variables with an identical experimental setup and a similar sample in both test environments. The influence of the test environment (independent variable) on the measures can then be determined by differences regarding the dependent variables. A comparison of the measures on the test environment makes it possible to determine the absolute validity of the static driving simulator. Within each test environment, braking intensity and timing of the acoustical warning as well as the driver’s annual kilometres, divided into two groups (occasional and frequent drivers) are the independent variables. The relative validity is examined by the influence of the independent variables on the measures within one test environment and compared to the other.

Results

Example for analysis procedure

First of all, hypotheses were literature-based formulated and then tested using t-tests, ANOVAs or correlations for all relevant scenarios regarding the measure under investigation. In total 52 hypotheses were formulated for objective and 46 for subjective measures. Furthermore 27 hypotheses focused on the relation between the measures. The analysis procedure will be explained using two of the five hypotheses concerning the objective measure speed:

_**H1**: Participants will choose lower speed when driving a full braking type of the AEBS, compared to other braking intensity types, in scenario 1. (explained by driver’s compensation of shorter following distance when distracted)
**H2**: The measure speed does not show differences comparing the two experimental environments for scenario 1, 3 and the 1st run of scenario 5.

The recorded values for speed at the beginning of the manoeuvre for the three braking intensities are shown in table 3.

### Table 3. Speed at the beginning of the manoeuvre for all braking intensities

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Speed [km/h] at beginning of maneuver</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full braking</td>
</tr>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Scenario 1</td>
<td>Test track</td>
</tr>
<tr>
<td></td>
<td>Driving sim</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>Test track</td>
</tr>
<tr>
<td></td>
<td>Driving sim</td>
</tr>
<tr>
<td>Scenario 5 (1st run)</td>
<td>Test track</td>
</tr>
<tr>
<td></td>
<td>Driving sim</td>
</tr>
</tbody>
</table>

Neither on the test track (ANOVA, $F(2,72) = 1.555; p = .218$) nor in the driving simulator (ANOVA, $F(2,80) = 2.883; p = .062$) data regarding speed at the beginning of the manoeuvre show differences depending on the braking intensity in scenario 1. **H1** is rejected in both experimental environments. Participants are generally driving 3 to 5 km/h faster in the driving simulator than on the test track. *t*-test are proving significant speed differences between simulator and test track for all braking intensities (Full Braking: $t(51) = -5.699; p < .001$; Partial Braking: $t(48) = -8.220; p < .001$; Combined braking: $t(55) = -4.061; p < .001$). **H2** is also rejected.

While relative validity can be confirmed, absolute validity is not given at first sight due to higher values in the driving simulator. Taking into account that speedometers in real vehicles are always indicating a velocity that is higher than the one that is actually driven, this result is not very surprising. For the objective measure speed it can be concluded that speedometers in the driving simulator must use an offset, similar to real vehicles, to gain data in the driving simulator that is showing absolute validity. Taking this into consideration when designing experiments in driving simulators speed can be seen as a suitable measure for evaluating ADAS.

**Interpretation of results and conclusion**

As a result of the study, insights were gained on the suitability of objective and subjective measures for evaluating ADAS intervening into the longitudinal control of the vehicle in a static driving simulator. The analysis for all measures was similar to the described procedure above. Measures can be distinguished between the ones which are suitable, partly suitable and not suitable. Suitable measures are showing mostly relative and absolute validity or the differences found between the experimental environments can be minimized simply, as shown on the example speed above. Partly suitable measures are showing in most scenarios relative or absolute validity while not suitable measure are mostly not showing relative nor
is simulation (not) enough?

In Table 4 suitable measures are marked in green, partly suitable in yellow and not suitable in red.

Table 4. Summary of suitability of the measures

<table>
<thead>
<tr>
<th>Objective measures</th>
<th>Subjective measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vehicle dynamics</strong></td>
<td></td>
</tr>
<tr>
<td>Speed</td>
<td>Acceptance</td>
</tr>
<tr>
<td><strong>Longitudinal acceleration</strong></td>
<td>Attitude toward the behaviour</td>
</tr>
<tr>
<td>Distance</td>
<td>Intention (to use the ADAS)</td>
</tr>
<tr>
<td>THW and TTC at driver’s intervention</td>
<td>Perceived system’s characteristics</td>
</tr>
<tr>
<td>Minimum and average THW</td>
<td>System evaluation</td>
</tr>
<tr>
<td><strong>Brake reaction time</strong></td>
<td>Moment and usefulness of intervention /</td>
</tr>
<tr>
<td></td>
<td>acoustical warning and best variant</td>
</tr>
<tr>
<td>Pedal Measures</td>
<td>Brake intensity and overall satisfaction</td>
</tr>
<tr>
<td>Movement time and number of driver’s intervention in perilous situations</td>
<td>Situation validation</td>
</tr>
<tr>
<td>Maximum of brake pedal actuation and time to reach maximum</td>
<td></td>
</tr>
<tr>
<td>Pedal activity in non-perilous situations</td>
<td>Probability of accident without the ADAS</td>
</tr>
<tr>
<td></td>
<td>and evading distance</td>
</tr>
<tr>
<td><strong>Steering behaviour</strong></td>
<td></td>
</tr>
<tr>
<td>Steering wheel angle</td>
<td></td>
</tr>
<tr>
<td>Evading reaction</td>
<td></td>
</tr>
<tr>
<td><strong>Driver behaviour</strong></td>
<td></td>
</tr>
<tr>
<td>Glance behaviour</td>
<td>suitable, partly suitable, not suitable</td>
</tr>
</tbody>
</table>

The results show that an initial assessment of intervening ADAS in a static driving is possible since subjective measures are mostly suitable or partly suitable. The participants react to imminent collisions in longitudinal traffic in a driving simulator similar to real vehicles. However, especially on the objective measures there are non-negligible differences. Measured longitudinal acceleration in the driving simulator within the first 0.5 s after braking intervention is generally lower than on the test track. This can be explained by higher brake reaction times in the driving simulator. Participants are also showing higher minimum and average THW when following a leading vehicle in the driving simulator (scenario 2). In the driving simulator experiment participants are not showing evading reactions during the braking manoeuvre which was frequently observed on the test track. To avoid misinterpretation, these restrictions should be strictly taken into consideration.

A subjective evaluation by the participants allows in the driving simulator a very good assessment of the system’s characteristics. The relations between the objective measures and the system and situation evaluation are identical to those on the test track. This implies that not only the results of the questionnaires are similar between the two experimental environments. Also their occurrence, in relation to the actual
behaviour of the participants, can be explained profoundly. There is a higher relation
between the overall acceptance and the system and situation evaluation on the test
track compared to the driving simulator. Even if acceptance in the driving simulator
is not significantly different from those measured on the test track, this can be seen
as an indication that the understanding of the system is more consistent on the test
track.

Discussion and outlook

An experiment has been designed and conducted on a test track and in a static
driving simulator to determine the (driving) behaviour during interventions of
system characteristics of an AEBS. The system was experienced and evaluated by
80 participants in each experimental environment. The focus was laid upon chosen
objective and subjective measures in order to derive conclusions about their
suitability for experiments to evaluate longitudinal intervening ADAS in a static
driving simulator.

The experiments have been carried out with unbiased participants. This means that
they were unfamiliar with the AEBS. A self-braking vehicle calls out an enormous
enthusiasm at first glance. This could be the reason why different system
characteristics and possible disadvantages were not recognized by the participants
and the acceptance and system evaluation was very positive in both environments.
The described study showed that despite the missing haptic feedback in the static
driving simulator participants are able to give an evaluation of the system similar to
when they are experiencing the AEBS in a real vehicle. This allows incorporating
static driving simulators in an early stage of the development process.

Literature

Human Decision Processes, 50*, 179-211.
Fahrassistenzsystemen. In S. Schade, and A. Engeln, A. (Eds.), *Fortschritte der
Verkehrspychologie* (pp. 313-337). Wiesbaden: GWV Fachverlage.
Studies.
Sicherheit? In *VDI-Bericht Nr. 1768* (pp. 25-44). Düsseldorf: VDI-Verlag.
Bryant, F.B. (2000). Assessing the Validity of Measurement. In L.G. Grimm and
P.R. Yarnold (Eds.). *Reading and understanding MORE multivariate statistics*
Realversuchen am Beispiel von Blickverhalten und Aktiver Gefahrenbremsung.
is simulation (not) enough?


